

AD A123185

12

RECOVERING INTRINSIC SCENE CHARACTERISTICS FROM IMAGES

Final Technical Report
Covering the Period 28 September 1981 to 27 September 1982

November 1982

Contract Amount:	\$415,508
Effective Date:	September 28, 1979
Expiration Date:	September 27, 1982

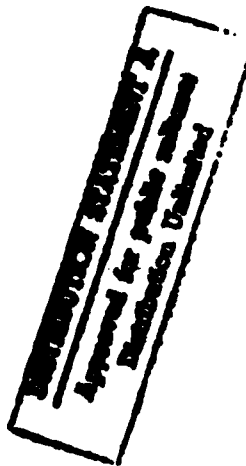
By: Martin A. Fischler, Program Director
Artificial Intelligence Center
Computer Science and Technology Division

Prepared for:
U.S. Army Research Office
P. O. Box 12211
Research Triangle Park, North Carolina 27709
Attention: Information Processing Office

Sponsored by:
Defense Advanced Research Projects Agency (DoD)
ARPA Order No. 2894
Monitored by U.S. Army Research Office
Under Contract No. DAAG29-79-C-0216

Contract No. DAAG29-79-C-0216
SRI Project 1019

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025
(415) 326-6200
Cable: SRI INTL MPK
TWX: 910-373-2046



83 01 10 016

DTIC FILE COPY



SRI International



RECOVERING INTRINSIC SCENE CHARACTERISTICS FROM IMAGES

Final Technical Report
Covering the Period 28 September 1981 to 27 September 1982

November 1982

Contract Amount:	\$415,508
Effective Date:	September 28, 1979
Expiration Date:	September 27, 1982

By: Martin A. Fischler, Program Director
Artificial Intelligence Center
Computer Science and Technology Division

Prepared for:
U.S. Army Research Office
P. O. Box 12211
Research Triangle Park, North Carolina 27709
Attention: Information Processing Office

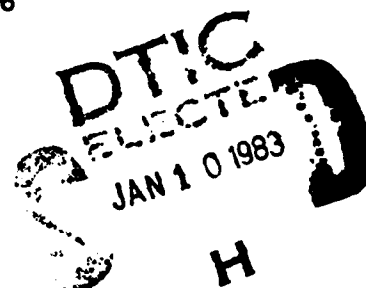
Sponsored by:
Defense Advanced Research Projects Agency (DoD)
ARPA Order No. 2894
Monitored by U.S. Army Research Office
Under Contract No. DAAG29-79-C-0216

Contract No. DAAG29-79-C-0216
SRI Project 1019

Approved:

Nils J. Nilsson, Director
Artificial Intelligence Center

David H. Brandin, Vice President and Director
Computer Science and Technology Division



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER Project 1019 Final Report	2. GOVT ACCESSION NO. AD-A223285	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) RECOVERING INTRINSIC SCENE CHARACTERISTICS FROM IMAGES		5. TYPE OF REPORT & PERIOD COVERED Final Report 9/28/79 through 11/27/82	
7. AUTHOR(s) Martin A. Fischler, Program Director Artificial Intelligence Center		6. PERFORMING ORG. REPORT NUMBER	
9. PERFORMING ORGANIZATION NAME AND ADDRESS SRI International Computer Science and Technology Division 333 Ravenswood Avenue, Menlo Park, CA 94025		8. CONTRACT OR GRANT NUMBER(s) DAAG29-79-C-0216	
11. CONTROLLING OFFICE NAME AND ADDRESS Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61101E	
14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office) DCASMA, San Francisco 1250 Bayhill Drive San Bruno, CA 94066		12. REPORT DATE 17 DEC 1982	13. NO. OF PAGES 96
		15. SECURITY CLASS. (of this report) Unclassified	
16. DISTRIBUTION STATEMENT (of this report) Approved for public release; distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Scene Analysis, Shape from Shading, Skyline Delineation, Edge Classification, Photometric Modeling, Imaging Constraint Models.			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) → The central theme of the research effort described in this report is the recovery of information about the three-dimensional structure and physical characteristics of the surfaces depicted in an image -- their shapes, locations, and photometric properties. Our approach has been to identify and model physically meaningful information that can either directly recover physical structure (e.g., models relating surface geometry to intensity variations visible in the image), or that can be used to constrain the interpretation process (e.g., camera and illumination models). (cont'd)			

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. KEY WORDS (Continued)

20 ABSTRACT (Continued)

→ In this report, we present new results concerning the question of how to recover the shape of a surface from its photometric variation in an image and a new method for skyline delineation. An essential feature of these techniques is that they are based entirely on physical models, and thus require no prior knowledge of the scene being analyzed. ←

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or
	Special
A	



DD FORM 1473 (BACK)
1 JAN 73

EDITION OF 1 NOV 66 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

I	INTRODUCTION	1
II	RESEARCH ACCOMPLISHMENTS	2
	REFERENCES	6

Appendix A--LOCAL SHADING ANALYSIS

Appendix B--FROM IMAGE IRRADIANCE TO SURFACE ORIENTATION

Appendix C--DETECTION OF THE VISIBLE SKYLINE
IN URBAN SCENES AND NATURAL TERRAIN

I INTRODUCTION

The central theme of our research is the recovery of information about the three-dimensional structure and physical characteristics of the surfaces depicted in an image--their shapes, locations, and photometric properties. The main obstacle to the recovery of such information is that images are inherently ambiguous representations of the scenes they depict. They are two-dimensional views of three-dimensional space, single slices in time of ongoing physical and semantic processes, and the light waves from which images are constructed carry limited information about the surfaces from which these waves are reflected. Consequently, interpretation cannot be based strictly on information contained in the image; it must also involve some combination of a priori models, constraints, and assumptions.

Our approach has been to identify and model physically meaningful information that can be used to constrain the interpretation process. The models we have developed fall into two categories: those that can be used to reconstruct and label scene content directly (e.g., models relating surface geometry and the physical structure of edges to the intensity variations visible in the image), and those models that identify the scene context and the conditions under which the image was produced (e.g., camera and illumination models, and models of the semantic content of the scene). The second class of models is more global in nature and is used primarily to constrain and control the application of the lower-level models in the first class.

The research personnel involved in this program include M. A. Fischler, J. M. Tenenbaum*, H. G. Barrow,* A. P. Witkin,* M. R. Lowry, G. Smith and A. P. Pentland.

II RESEARCH ACCOMPLISHMENTS

Much of our early work addressed the issue of recovering surface geometry (e.g., orientation and depth) and reflectance from the assumed continuity of the surfaces, knowledge of the nature of the edges bounding the surfaces, and the intensity variations measured in the image. Surface perception plays a fundamental role in early visual processing, both in humans and in machines. An explicit representation of surface structure is necessary for many low-level visual functions involved in such applications as terrain modeling, remote sensing, navigation, manipulation, and obstacle avoidance. It is also a prerequisite for general-purpose vision systems capable of human-level performance in such tasks as object recognition and scene description. References [1] and [2] describe our work on surface modeling.

Our efforts in the later stages of this project were concerned with the issues of edge modeling (i.e., edge classification), illumination and intensity modeling (i.e., the relationship between surface reflectance and image intensity), and modeling of the geometric

* Now employed at Fairchild Camera and Instrument Corporation, Palo Alto, California.

constraints introduced by the imaging process (camera model and vanishing points).

We have made significant progress in the following problem areas:
(A description of this work is presented in references [3] through [6] and Appendices A through C.)

- (1) The ability to identify the physical nature of an edge (e.g., occlusion edge, shadow edge, etc.) solely according to its photometric appearance in a black-and-white image. This capability is necessary for recovering surface orientation and shape from a single image on the basis of shading and texture variations [Ref 3].
- (2) The ability to quickly locate straight parallel edges characteristic of man-made structures (e.g., the edges of buildings) [Ref 3].
- (3) The ability to recover absolute scene intensity information without calibration data (such as a step wedge exposed on the image) based on knowing the identity of the material composition of the surfaces at a few points in the image. This capability is necessary for partitioning the image into labeled regions of a given material type [Ref 3].
- (4) The ability to identify the skyline in an image on the basis of simple models of the relationship among land, sky, and cloud brightness and texture [Appendix C].

In Appendices A and B we provide new results concerning the question of how to recover the shape of a surface from its appearance in an image.

The relationship between image intensity and the orientation of the surface responsible for that image is usually expressed in terms of the image irradiance equation. This equation requires that the

characteristics of the scene illumination and the surface material be known. The shape-from-shading task (recovering surface orientation from image irradiance) has amounted to solving this equation. Of the various methods that might be used to solve this equation, relaxation-style algorithms seem to offer the greatest potential when we are working with natural imagery. To recover surface orientation, relaxation-style algorithms based on the image-irradiance equation employ additional constraints. These constraints, which are needed to supplement the underdetermined image-irradiance equation, usually capture the concept of smoothness. While smoothness superficially determines the relationship between image irradiance and surface orientation, we have found that smoothness is too weak a concept to propagate boundary conditions and is thus equally ineffectual as a means of recovering the required solution.

In Appendix A we derive a new formulation for the shape-from-shading task; we have traded the need to know the explicit form of the scene-radiance function for the assumption that material scatters light isotropically. This model is applicable to natural scenery without additional assumptions about illumination conditions or the albedo of the surface material (the model also demonstrates some competence even when the scattering is not isotropic).

In Appendix B we address the question: What constraints on surface shape are imposed by local image data from a single image? We show that a local analysis of image shading, with no other knowledge about the

viewed scene, may be used to estimate surface shape. Specifically, we prove that, given a single image, the following conditions obtain:

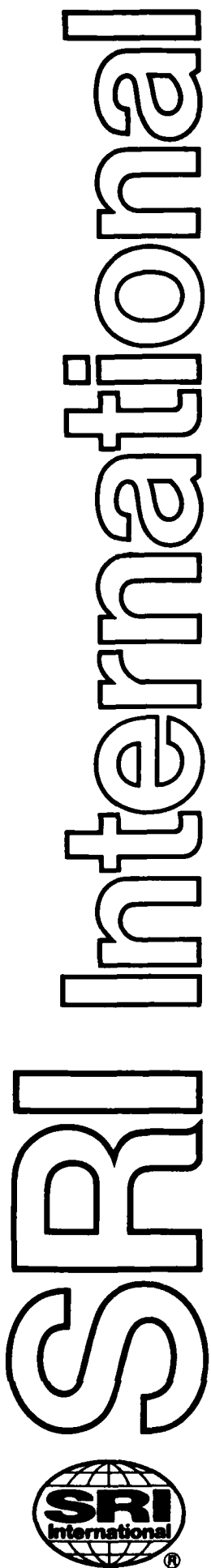
- (1) Surface orientation may be determined exactly at umbilical points (points with equal principal curvatures) on a Lambertian surface through a local analysis of the image without knowledge of illumination, surface curvature, albedo, or boundary conditions. Recovery of surface orientation allows local determination of illuminant direction, surface curvature, and the product of illuminant intensity and surface albedo.
- (2) The image intensity, together with its first and second derivatives at each image point, are identical to those of an imaged Lambertian umbilical point with unique orientation, curvature, (overhead) illumination and product of albedo times illumination intensity. This implies that it is impossible for any local analysis to determine whether or not (a) a surface is Lambertian, and (b) the principal curvatures are equal.

When a person views a scene, he has an appreciation of where he is relative to the scene, which way is up, the general geometric configuration of the surfaces (especially the support and barrier surfaces), and the overall semantic context of the scene. The research results we have described can provide similar information to constrain the more detailed interpretation requirements of machine vision (e.g., such tasks as stereo compilation and image matching).

REFERENCES

1. H.G. Barrow and J.M. Tenenbaum, "Interpreting Line Drawings as Three-Dimensional Surfaces," Artificial Intelligence, Vol. 17(1), pp. 75-116 (August 1981).
2. A.P. Witkin, "Recovering Surface Shape and Orientation from Texture," Artificial Intelligence, Vol. 17(1), pp. 17-45 (August 1981).
3. M.A. Fischler, S.T. Barnard, R.C. Bolles, M. Lo, L. Quam, G. Smith, A.P. Witkin, "Modeling and Using Physical Constraints in Scene Analysis," Artificial Intelligence Center Technical Report, SRI International, Menlo Park, California (April 1982). Also published in Proceedings of the National Conference on Artificial Intelligence, pp. 30-35 (August 1982).
4. G. Smith, "From Image Irradiance to Surface Orientation," Artificial Intelligence Center, Technical Note 273, SRI International, Menlo Park, California (December 1982) (Appendix A).
5. A.P. Pentland, "Local Shading Analysis," Artificial Intelligence Center Technical Note 272, SRI International, Menlo Park, California (November 1982) (Appendix B).
6. M.R. Lowry, "Detection of the Visible Skyline in Urban Scenes and Natural Terrain," SRI International, Menlo Park, California (December 1982) (Appendix C).

Appendix A
LOCAL SHADING ANALYSIS
Alex P. Pentland



LOCAL SHADING ANALYSIS

Technical Note 272

November 1982

By: Alex Paul Pentland, Computer Scientist

**Artificial Intelligence Center
Computer Science and Technology Division**

SRI Project 1019

The final stages of the research reported herein was supported at SRI by Army Research Office Contract DAAG29-79-C-0216. A substantial portion of this research was accomplished at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract N00014-80-C-0505, in part by National Science Foundation Grant 79-23110MCS.

ABSTRACT

Local analysis of image shading in the absence of prior knowledge about the viewed scene, may be used to provide information about the scene. The following has been proved.

Every image point has the same image intensity and first and second derivatives as the image of an umbilical point (a point with equal principal curvatures) on a Lambertian surface; there is *exactly one* combination of surface orientation, curvature, (overhead) illumination direction and albedo times illumination intensity that will produce a particular set of image intensity and first and second derivatives. A solution for the unique combination of surface orientation, etc., at umbilical points is presented

This solution has been extended by using general position and regional constraints to obtain estimates of the following:

- Surface orientation at each image point
- Whether the surface is planar, singly or doubly curved at each point
- The mean illuminant direction within a region
- Whether a region is convex, concave, or is a saddle surface.

Algorithms to recover illuminant direction, identify discontinuities, and estimate surface orientation have been evaluated on both natural and synthesized images, and have been found to produce useful information about the scene.

I. Introduction

A spatially restricted analysis of an image is logically the first stage of any visual system. This initial stage of analysis is especially important because it determines what information will be available to the remainder of the visual system; if a rich description of the world can be computed locally, there is a smaller computational load placed on the remainder of the system. It is, therefore, important to ascertain as much about the world as possible at this first stage of processing.

Biological visual systems conform to this principal. There is overwhelming evidence that they devote a large percentage of their neurons to an initial *local* analysis of the image. Thus, assessment of the limits and potential uses of a local analysis can be expected to provide insight into both machine and biological vision problems.

What information is available locally? When we examine a small neighborhood around an image point, we often find only small changes in shading (changes in image intensity¹). It is unusual to find a contour passing through an image point. Thus, if we are to learn about scene characteristics from local examination of an image, we must concern ourselves with shading². The main question posed in this paper will therefore be: What information can, or cannot, be recovered from an unfamiliar image through a local analysis of shading? In the following sections I shall first discuss the limitations that are inherent in any local analysis of image shading, and then show how information about the scene can be determined by means of additional constraints derived from general position and the distribution of data within homogeneous image regions. Proofs of the various propositions are presented in the appendix.

Previous work. Horn and his colleagues [2], [3] have analyzed the process of image formation and have developed several numerical integration schemes for using image intensity to solve for object shape. These shape-from-shading techniques, however, require considerable *a priori* knowledge of the scene, and they function by propagating constraint from boundary conditions (such as those provided by smooth occluding contours) over the surface whose shape is to be estimated. These techniques, therefore, cannot be applied to an unfamiliar, unanalyzed scene and do not perform the purely local analysis I wish to consider here.

Bruss [4] has addressed the question of whether shape can be derived from shading using a purely local analysis (again with a considerable *a priori* knowledge of the scene assumed). She proved that no shape-from-shading technique can yield a unique solution without additional constraints which, in certain restricted cases (most importantly electron

¹To avoid confusion, the term "image intensity" will be used throughout this document, rather than the technically more correct "image irradiance," both for the flux per unit area falling on the image plane and for the measured image irradiance. The two may be assumed to be numerically equal; and thus the distinction has little significance for the task at hand.

²For the purposes of this paper we shall restrict our attention to shading, because the problem of estimating shape from local texture information has already received much attention, e.g., [17] and [18].

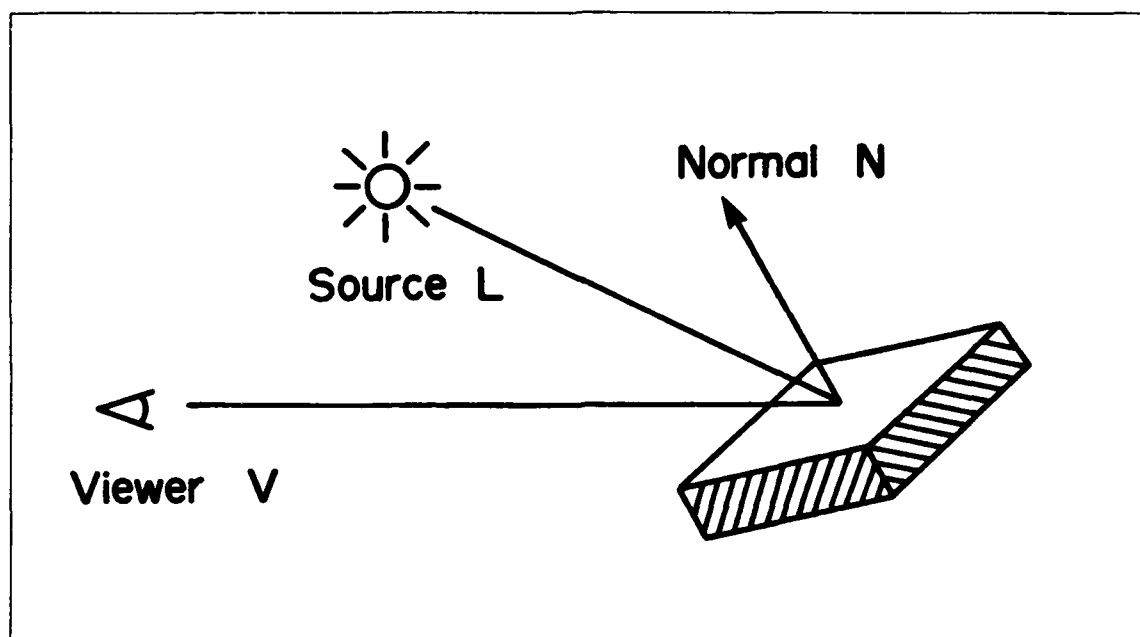


Figure 1. A simple model of image generation. N is the surface normal, L the illumination direction, V the viewer's direction. If λ is the flux emitted toward the surface, ρ the average reflectance of the surface, and we assume distant light source and a Lambertian reflectance function for the surface, then the image intensity I is given by $I = \rho\lambda(N \cdot L)$.

micrographs), may be provided by the bounding contour of the surface. Bruss, however, dealt mostly with the question of what *cannot* be obtained from an analysis of shading; the question of what *can* be accomplished with a local analysis of shading was not fully explored. It is this, consequently, that we discussed below.

A. Image Formation

Before we can make quantitative statements about the limitations or usefulness of a local analysis of shading, we must first develop a mathematical model of the image generation process. Figure 1 shows a simple model of image generation: a distant point-source illuminant at direction L , a patch of surface with surface normal N , and a viewer in direction³ V . We will assume orthographic projection; note, however, because the model is purely local orthographic and perspective projection are identical except at points of discontinuity.

The surface normal N , the viewer's direction V and the illuminant direction L are unit vectors in Cartesian three-space. As they are unit vectors, two parameters suffice to specify them, the third being determined by the constraint that they have unit magnitude. Two parameters that are often chosen are the slant σ and the tilt τ . The tilt of a surface is the image-plane component of surface orientation and is equal to $\tan^{-1}(y_N/x_N)$, where x_N and y_N are the x and y components of the surface normal. The slant of the surface is the depth component of surface orientation and is equal to $\cos^{-1}(z_N)$, where z_N is the z

³All boldface variables (e.g., N , L , p etc.) represent three-dimensional vectors (x, y, z) , all other variables are scalars. The (x, y) plane is taken parallel to the image plane, so that $V = (0, 0, 1)$.

component of the surface normal.

The image intensity I is in general given by

$$I = \rho \lambda (\mathbf{N} \cdot \mathbf{L}) R(\mathbf{N}, \mathbf{L}, \mathbf{V}) (\mathbf{N} \cdot \mathbf{V})^{-1}$$

where ρ , the *albedo*, is the portion of incident light that is reflected, λ is the amount of light incident upon the surface and $R(\mathbf{N}, \mathbf{L}, \mathbf{V})$ is the reflectance function, which describes how much of the reflected light leaves in each direction. The amount of incident light reflected in the viewer's direction \mathbf{V} is a function of the illuminant direction \mathbf{L} and the surface normal \mathbf{N} . The term $(\mathbf{N} \cdot \mathbf{L})$ describes the amount of light incident upon the surface, while the term $(\mathbf{N} \cdot \mathbf{V})^{-1}$ describes the foreshortening that occurs during projection into the image⁴. A Lambertian reflectance function, an idealization of rough, matte surfaces, is defined as

$$R(\mathbf{N}, \mathbf{L}, \mathbf{V}) = \mathbf{N} \cdot \mathbf{V}$$

which is proportional to the reciprocal of the foreshortening caused by the projection term. Thus, for a Lambertian surface the reflectance function and the effect of projection cancel each other, and the equation for image intensity becomes

$$I = \rho \lambda (\mathbf{N} \cdot \mathbf{L}) \quad (1)$$

Thus, the assumption of a Lambertian reflectance function is equivalent to the assumption that the scattering of incident light is isotropic. We shall assume a Lambertian reflectance function.

Generality of the assumptions. The assumption of a distant point-source illuminant and a Lambertian reflectance function is not as restrictive as it might at first seem. We note, for instance, that for a Lambertian surface *any* constant distribution of illumination is equivalent to a single distant point-source illuminant; this follows from application of the mean value theorem. Because we are concerned only with local analysis, the requirement that the distribution of illumination be constant is almost trivially met⁵. Therefore, local inferences derived with this single-illuminant/ Lambertian model will generally be valid whenever the surface scatters incident light in an isotropic manner, regardless of the actual distribution of illumination.

B. The Derivatives Of Image Intensity

The image intensity I and the surface normal \mathbf{N} are different at each point (x, y) in the image, and thus are perhaps better written $I(x, y)$ and $\mathbf{N}(x, y)$. However, when they are discussed at a particular point P , they will be written as simply I and \mathbf{N} . Similarly, we shall write dI and $d\mathbf{N}$ to designate the first derivative of image intensity and the surface normal, respectively, at a point P in the direction (dx, dy) . The partials of I , \mathbf{N} , and other variables will be denoted by subscripts, i.e., $I_x = \partial I / \partial x$ and $\mathbf{N}_y = \partial \mathbf{N} / \partial y$.

⁴In other terminology, $\mathbf{N} \cdot \mathbf{L}$ is equal to the cosine of the incident angle, and $\mathbf{N} \cdot \mathbf{V}$ is equal to the cosine of the exitent angle.

⁵Only illumination near the "horizon" of the surface patch causes a problem; in this case there is some self-occlusion and thus somewhat different illumination at neighboring points.

If we are examining a small, homogeneous region of an image, it is reasonable to assume that the illumination and albedo of the surface change very little, and so we may treat L , ρ , and λ as constants. If we also assume a Lambertian reflectance function, so that Equation (1) applies, then

$$dI = d(\rho\lambda(N \cdot L)) = \rho\lambda(dN \cdot L) + \rho\lambda(N \cdot dL) = \rho\lambda(dN \cdot L) \quad (2)$$

The term $(N \cdot dL)$ is zero because L was assumed constant. Similarly, the second derivative of image intensity is

$$d^2I = d(\rho\lambda(dN \cdot L)) = \rho\lambda(d^2N \cdot L) + \rho\lambda(dN \cdot dL) = \rho\lambda(d^2N \cdot L) \quad (3)$$

Thus, the second derivative of image intensity depends upon the second derivative of the surface normal, just as the first derivative depended upon the first derivative of the surface normal.

II. Local Shading Information: Limitations And Potential

Before we can know what is possible to accomplish with local shading information, it is important to characterize what *cannot* be done. The following proposition describes the fundamental limitation which is inherent to any local analysis of image shading:

Proposition 1. The image of a Lambertian umbilical point (a point with equal principal curvatures) can produce any combination of image intensity I and derivatives I_x , I_y , I_{xx} , I_{yy} and I_{xy} .

This proposition says that when we view a point on a surface, regardless of what the actual surface curvatures are or what the actual surface reflectance function is, the resulting image point always looks like an umbilical point on a Lambertian surface. This proposition implies, therefore, that it is impossible for a local analysis of the image to determine unambiguously whether a surface is Lambertian and whether the principal curvatures are equal; there will always be the possibility that the observed point is an umbilical point on a Lambertian surface. We cannot resolve these ambiguities by resorting to higher derivatives of image intensity because, although more measurements are obtained by measuring the higher derivatives, each additional derivative brings in more unknowns than measurements.

We can see by the following argument that this proposition is likely to be true. Consider that at each point in an image we can measure the intensity, and its first and second derivatives to obtain six independent measurements, which are I , I_x , I_y , I_{xx} , I_{yy} and I_{xy} . To specify the image intensity of an umbilical point on a Lambertian surface⁶ requires six independent parameters: τ the surface tilt⁷, σ the surface slant⁸, R the radius of curvature, $(l_1, l_2, \sqrt{1 - l_1^2 - l_2^2})$ the illuminant direction, and $\rho\lambda$ the surface albedo times illuminant intensity. Solving for these six unknowns requires at least one measurement for

⁶The set of all possible images of umbilical points is obtained by considering surfaces of the form $z(x, y) = \sqrt{R^2 - x^2 - y^2}$ for particular values of $R > 0$, $R \geq x \geq -R$, $R \geq y \geq -R$.

⁷Tilt is the image-plane component of surface orientation.

⁸Slant is the depth-component of surface orientation.

each unknown; thus, the measurement of intensity, first and second derivatives can at most establish the six parameters required to specify a Lambertian umbilical point. No additional measurements are available to determine whether the surface curvatures are unequal or whether the reflectance function is Lambertian.

Proposition 1 leaves open the possibility that there may be a great many combinations of surface orientation, curvature, etc., corresponding to each combination of image intensity and its derivatives. If the equations for image intensity were linear, there would be exactly one combination of the six parameters that would correspond to the observed measurements. Although the equations are not linear, the following proposition shows that there are only two possible combinations of these factors that will yield a particular combination of image intensity and derivatives.

Proposition 2. Given the image of an umbilical point on a Lambertian surface with image intensity I and derivatives I_x , I_y , I_{xx} , I_{yy} and I_{xy} there are two possible combinations of surface orientation, curvature, illuminant direction and surface albedo times illuminant intensity, one with the illuminant direction above the line of sight, the other exactly opposite surface tilt and illuminant tilt.

Thus, the ambiguity present in local image shading is not much greater than was evident from the first proposition; there is only the additional ambiguity that arises from a symmetry involving the illuminant direction and the tilt of the surface. This symmetry results in the interrelation of the direction of illumination and the convexity of the surface; if the illuminant direction is taken to have the opposite tilt (e.g., from above the line of sight to below it) the convexity of the surface will reverse. Therefore, one cannot determine the convexity of the surface unless something is known about the illuminant direction [1], [5].

Using propositions 1 and 2, we can produce an exhaustive characterization of the limitations of any local analysis of shading. A local analysis of shading cannot

- Determine the sign and magnitude of the surface curvatures⁹ e.g., whether the surface is convex, concave or a saddle-shaped and whether or not the curvatures are unequal.
- Determine the surface reflectance function.
- Separate the surface albedo from illuminant intensity.

A. Solving For Image Formation Parameters At An Umbilical Point

The amount of information we can extract from a local analysis of shading (given that we are viewing the image of a Lambertian umbilical point) is surprising. For such image points we can solve for *every* parameter in the image formation process. This seems to indicate that there is approximately two degrees of freedom left undetermined by the local shading information when we view a point within a homogeneous region of an image: the ratio of the surface curvatures and the degree to which the surface is non-Lambertian.

The umbilical-point case is the most complex situation in which all of the image formation parameters may be recovered locally. The fact that there are only relatively

⁹With the additional constraint provided by general position, we can determine when the curvatures are zero. This is shown in the following section.

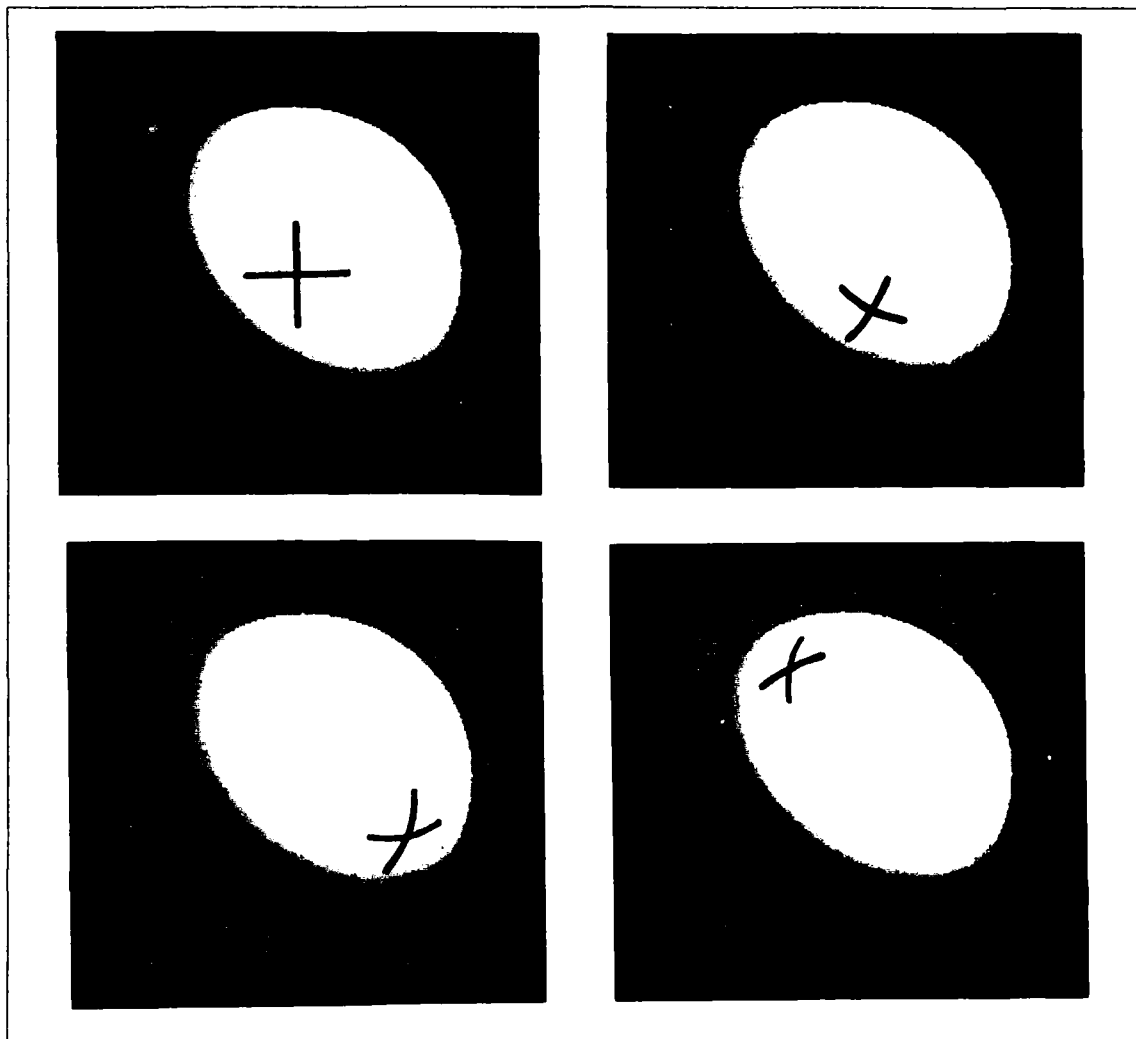


Figure 2. The manner in which image curvature "spreads" indicates the tilt of the surface. This may be understood by imagining that we could observe the lines of curvature on a surface directly. They would look just like the lines drawn in this figure. If we were looking straight down on the surface of a sphere, the lines of curvature would appear perpendicular, as in (a). As we tilted the surface to one side, the lines of curvature would appear progressively more spread, as in (b) and (c). Different directions of tilt cause spreading in different directions, as demonstrated in (d). The amount of spread depends on the slant of the surface.

few additional parameters required to obtain a reasonably general model suggests that the umbilical-point solution may provide us with a useful (albeit simplified) model of how the various portions of the image formation process evidence themselves in the image, and may also prove useful as a tool for analyzing image points. The umbilical-point solution for surface orientation, for example, is instructive to examine. How can surface orientation be determined from local shading information?

Imagine that we could observe the lines of curvature on a surface directly. They would look like the lines drawn in Figure 2. If we were looking straight down on the surface of a sphere, the lines of curvature would appear perpendicular, as in Figure 2 (a). As we tilted the surface to one side, the lines of curvature would appear progressively more spread, as in

Figures 2 (b) and (c). Different directions of tilt would cause spreading in different directions, as demonstrated in (d).

We cannot observe lines of curvature on the surface directly, of course, but we can observe the interaction of surface curvature with the illuminant in the second derivatives of image intensity. The second derivative of image intensity has three components: I_{xx} and I_{yy} , the "curvature" of image intensity along the x and y axes, respectively, and I_{xy} , the "spread" of those curvatures. Just as with the spread of the lines of curvature, the direction in which the spread term is greatest is also the direction of the surface tilt. The direction in which this spread is greatest also turns out to be the direction along which d^2I is greatest, and hence the following proposition:

Proposition 3. Given the image of an umbilical point on a Lambertian surface, the tilt of the surface τ is the image direction in which the second derivative of image intensity d^2I is greatest.

Thus, for Lambertian umbilical points the tilt may be determined from the second derivative of image intensity directly, without *a priori* knowledge. This leaves only the surface slant to be determined.

In Figure 2 the direction of the spread indicated the tilt of the surface. Similarly, the *amount* of the spread indicates the slant (depth) component of the surface orientation. Measuring the magnitude of this spread relative to the total curvature (as measured by the Laplacian) provides an indicator of the surface slant, as described in the following proposition.

Proposition 4. Given the image of an umbilical point of a Lambertian surface, the surface slant σ is given by

$$\sigma = \cos^{-1} \sqrt{\frac{k \nabla^2 I - (k^2 + 1) I_{xy}}{k \nabla^2 I + (k^2 + 1) I_{xy}}}$$

where $k = \tan^{-1} \tau$.

The slant and the tilt propositions together determine surface orientation exactly. Note that neither the slant nor the tilt estimate requires *any* knowledge of illuminant direction, surface albedo, curvature, or illuminant intensity.

B. Unlikely Umbilical-Point Solutions: Constraint From General Position

Although the umbilical-point solution will always provide us with an interpretation that is consistent with the local image data, it sometimes yields an interpretation of an image point that strikes us as *unlikely* because the umbilical-point solution requires an unlikely configuration of orientation, illumination, or viewer direction.

When we observe an image point for which the umbilical-point solution requires an unlikely configuration, we have available the additional constraint provided by general position to help us interpret the image data. This additional constraint allows us to reject the umbilical-point solution and infer that something special has occurred in the image formation process — something that may permit further analysis.

Zero second derivative. One such case arises when one or both of the second derivatives are zero. When one of the second derivatives is zero, the umbilical-point solution is a

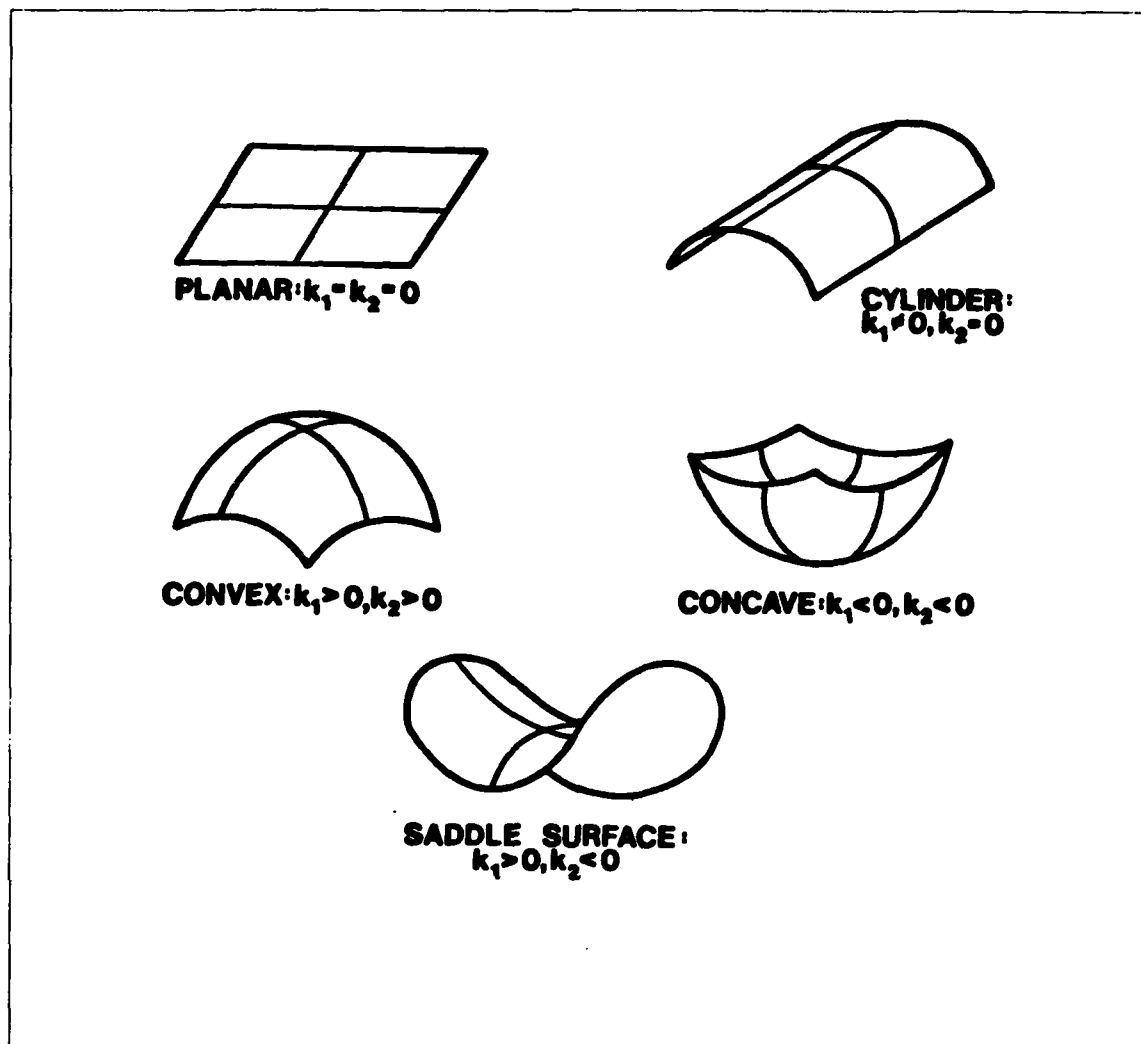


Figure 3. Surface types. Surfaces may be classified into five types: planar, cylindrical, convex, concave, or saddle surface. The classification of a surface depends on whether the two principal curvatures κ_1 and κ_2 are positive, negative, or zero.

surface patch whose orientation is exactly perpendicular to the line of sight. When both of the second derivatives are zero, the umbilical-point solution requires that the illuminant direction be exactly in the image plane. These interpretations of the image point are unlikely because precise alignment of surface orientation or of illuminant direction is necessary, i.e., this interpretation of the image point presupposes a violation of general position. The more likely inference when a zero second derivative is observed is that one of the surface curvatures is zero. The fact that this inference is valid (shown in the appendix) allows us to partially classify the surface type.

Surface points may be classified into five types: planar, cylindrical, convex, concave, or saddle surface. These five types are shown in figure 3. The classification of a surface depends on whether the two principal curvatures κ_1 and κ_2 are positive, negative, or zero:

<i>plane</i>	$\kappa_1 = 0$	$\kappa_2 = 0$
<i>cylinder</i>	$\kappa_1 \neq 0$	$\kappa_2 = 0$
<i>convex</i>	$\kappa_1 < 0$	$\kappa_2 < 0$
<i>concave</i>	$\kappa_1 > 0$	$\kappa_2 > 0$
<i>saddle surface</i>	$\kappa_1 > 0$	$\kappa_2 < 0$

One important step in identifying the type of surface is determining when the principal curvatures are zero, as this allows us to classify the surface as planar, cylindrical, or doubly curved.

When one of the principal curvatures is zero, the surface normal does not change as we travel along the surface in the direction of that principal curvature. Because the surface normal N does not change along that direction (let us specify the direction by (dx, dy)), we know that¹⁰ $dN = 0$ along (dx, dy) . Since

$$dI = \rho \lambda dN \cdot L \quad (2)$$

we see that dI must also be zero along (dx, dy) . Unfortunately, the reverse inference is not generally true, because dI is zero along *some* direction for every image point. Therefore, we cannot infer that $dN = 0$ in direction (dx, dy) just because $dI = 0$ along that direction.

That problem does not occur when we observe $d^2I = 0$ along a direction (dx, dy) . When the surface normal does not change along a direction (dx, dy) , then $d^2N = 0$ along (dx, dy) . Since

$$d^2I = \rho \lambda d^2N \cdot L \quad (3)$$

we see that, when $d^2N = 0$ along (dx, dy) , then d^2I must also be zero along (dx, dy) . For the second derivative, the reverse inference — that $d^2N = 0$ because $d^2I = 0$ — is generally valid, for when we observe that $d^2I = 0$ along direction (dx, dy) , we can conclude that either (1) d^2N is perpendicular to L or (2) that $d^2N = 0$. As it is unlikely that d^2N is perpendicular to L for any distance, we may legitimately conclude that, if we observe that $d^2I = 0$ for some distance along direction (dx, dy) , then $d^2N = 0$. This implies that dN is constant along (dx, dy) , and, if dN remains constant for some distance, we may use the constraint of general position to conclude that N is also constant (this is shown in the appendix).

We can now begin to classify the surface. If we observe that $d^2I = 0$ along a line in the image, then N does not change along that locus and we have a surface that is cylindrical along that line. If we observe that $d^2I = 0$ along a direction (dx, dy) throughout some region in the image, then the surface is a cylinder with an axis whose projection points in the (dx, dy) direction. Similarly, if we observe that $d^2I = 0$ along two orthogonal directions, then N does not change along either direction and thus the surface is planar. Finally, if $d^2I \neq 0$ in all directions, then the surface must be doubly curved, i.e., it is convex, concave, or a saddle surface. We can not distinguish among these alternatives on the basis of local shading information alone — a consequence of the previous propositions. Thus, we have the following proposition:

¹⁰Here the '0' in ' $dN = 0$ ' is the zero vector.

Proposition 5. The surface type at a point is partially determined by the number of directions in which $d^2I = 0$.

$d^2I = 0$ in no directions \Rightarrow *convex/concave/saddle surface*

$d^2I = 0$ in one direction \Rightarrow *cylinder*

$d^2I = 0$ in all directions \Rightarrow *plane*

It is interesting to note that linear intensity gradients do not invalidate this classification scheme.

The detection of lines along which a surface is cylindrical is of considerable importance because it is only at such cylindrical lines that changes occur in the surface type (e.g., change from a convex to a saddle surface). As the surface changes from one type to another, the sign of at least one of the principal curvatures changes from positive to negative, or vice versa. In the course of a sign change the curvature is briefly zero, and so the surface is cylindrical along the locus where the surface changes type¹¹ Thus, lines along which $d^2I = 0$ are places where the surface is undergoing a change of type, and the set of such lines divides the surface into regions that are of the same surface type.

III. Generalization Of The Results: Regional Constraints

In real images, relatively few points are umbilical and relatively few surfaces are Lambertian. Therefore, we must find some additional constraints in order to obtain generally applicable formulas for surface orientation, illuminant direction, and so forth. Unfortunately, the thrust of the preceding propositions is that there is no point-wise local assumption that will generally be true; there will always be at least a two-parameter family of possible solutions.

One way we can obtain additional constraint is to expand our view: to consider *regions* rather than single points only. Once we allow discussion of regions, we find that there are many possibilities for obtaining a good estimate of the mean value of particular parameters within the region, by using inferences about the range or distribution of image data within the region. Having obtained an estimate for the mean value of a parameter, we can then solve for other parameters by *assuming* the already estimated value — i.e., by bootstrapping.

The mean value of a parameter within a region may be used either to comment upon the *average* properties of the region, or we may assume that the parameter is constant throughout the region and thus obtain point-by-point estimates. If we comment only about average properties, then the validity of our deductions depends solely on the accuracy of the initial estimate. If we desire point-by-point estimates, the validity of our inferences is also conditional upon the intraregional variance of the estimated parameter.

In the remainder of this paper I shall discuss results obtained by estimating the mean value of one parameter within a region and then using this estimate as an assumption to infer other properties of the scene. Examples of both regional and point-by-point inference will be presented. Estimates of average properties of a region, such as illumination direction and surface type (e.g., convex, concave, saddle, etc.), have been made by using the

¹¹If the change takes place over an extended area, both curvatures will be zero and so the surface will be planar instead of cylindric.

maximum-likelihood estimate of change in surface normal. Point-by-point estimates of surface orientation have been made by using an estimate of the curvature within the region.

A. Finding The Illuminant Direction

Estimating the illuminant direction is difficult because image data are determined by both the surface normal and the illuminant direction. Since evidence relating to illuminant direction is confounded with the unknown direction of the surface normal, estimating the direction of illumination seems to require making some assumption about surface orientation or its derivatives.

One useful assumption is that change in surface orientation ($d\mathbf{N}$) is distributed isotropically within each image region. It is true that $d\mathbf{N}$ is isotropically distributed when considered over all scenes; furthermore, there is a large class of common image regions for which $d\mathbf{N}$ is isotropically distributed. This class of image regions includes all images of convex objects bounded entirely by a gradual occluding contour¹², such as the image of a smooth pebble.

Given the assumption that changes in surface orientation are isotropically distributed, we can devise a procedure for estimating the illuminant direction \mathbf{L} by looking for the regular biasing effect of the illuminant direction on $d\bar{I}$, the mean value of dI , along various image directions (dx, dy). The effect of the illuminant direction is to make $d\bar{I}$, vary according to

$$\begin{aligned} d\bar{I} &= \rho \lambda d\bar{\mathbf{N}} \cdot \mathbf{L} \\ &= \rho \lambda (d\bar{x}_N x_L + d\bar{y}_N y_L + d\bar{z}_N z_L) \end{aligned}$$

where $d\bar{\mathbf{N}} = (d\bar{x}_N, d\bar{y}_N, d\bar{z}_N)$ is the mean change in $d\mathbf{N}$ measured in image direction (dx, dy), and $\mathbf{L} = (x_L, y_L, z_L)$ is the illuminant direction. Under the assumption that change in surface normal is distributed isotropically within a region, then, along any *one* image direction (dx, dy) we find that $d\bar{x}_N$ is proportional to dx , the x -component of the image direction, that $d\bar{y}_N$ is proportional to dy , the y -component of the image direction, and that $d\bar{z}_N$ is zero. (see [5]) Therefore,

$$d\bar{I} = k(x_L dx + y_L dy) \quad (4)$$

where k is a constant determined by the albedo, illuminant strength, and the variance of the distribution of $d\mathbf{N}$ within the region.

Using (4), we can set up a linear regression that employs the mean of dI as measured along various image directions to obtain a maximum-likelihood estimate of the ratio of the unknowns x_L and y_L . This ratio is the tilt of the illuminant direction, which we shall use in identifying surface type. The constant k (and from this the values of x_L, y_L and z_L) can be estimated from the mean and variance of the distribution of dI along various image directions. In this procedure, most of the information about the illuminant direction comes from image points where there are large changes in image intensity, e.g., edges and specularities. This seems to agree with our introspective impression as to how we determine

¹²This may be proved by noting that the surface normals on such an object are perpendicular to \mathbf{V} at the image boundary of such an object, and thus (given that the object is strictly convex) we may form a 1-1 onto map between the surface normals of the object and the Gaussian sphere, which has sum $d\mathbf{N}$ equal to zero.

the illuminant direction. Note that this estimation procedure establishes the tilt of the illuminant direction to within $\pm\pi/2$, leaving an ambiguity regarding illuminant position that is identical to the human perceptual ambiguity that obtains in the absence of cast shadow information.

Evaluation of the illuminant direction estimator. This illuminant estimation procedure has been compared with the answers given by fifteen human observers on a series of digitized pictures of natural objects, such as rocks and logs. The photographs of these objects were made in natural illumination, so that the imaged scenes do not have a Lambertian reflectance function or true point-source illumination. Digitized versions of the pictures were shown to the human subjects, so that both they and the computer procedure would receive exactly the same image information.

Figures 4 (a) and (b) show a comparison of human and computer estimates of illuminant direction. Previous experiments have documented that the fifteen subjects' mean estimates exhibit a standard error of ten degrees in this experimental condition. Thus, the human and computer estimates shown in Figure 4 concur to within experimental error.

Other evidence concerning the equivalence of human and computer estimates comes from the variance of the two estimates. The illuminant direction estimator generates a confidence statistic for each image, along with its estimate. This confidence statistic is proportional to the variance of the estimate for that image (given the assumptions of the procedure). We can compare the variance of human estimates for a particular picture with the variance of the maximum-likelihood estimate (as predicted by the confidence statistic). This comparison is shown in Figure 4 (c). There is a correlation of 0.63 between the variance of the two sets of estimates, significant at the $p = 0.05$ level. The linear regression line relating the human and maximum-likelihood variance is shown as a dashed line; the coefficients of the regression are significantly different from zero at the $p = 0.01$ level. The significant relationship between the variance of the two estimation procedures (human and computer) shows that, when one of them finds enough information in the image to make a low-variance estimate, so does its counterpart.

One of the images employed is of particular importance, because it is an example of incorrect estimation by humans of the illuminant direction. When the image of the rock shown in Figure 4 (d) was presented to human subjects, they misestimated the illuminant direction by about 120 degrees (it is actually illuminated from top left, not top right as all but two of the fifteen subjects reported). The computer generated estimate, interestingly, agrees with the human ones — even though in both cases the estimates are objectively wrong. This image must violate the assumptions on which the human estimates of illuminant direction are based, because the human estimate is objectively wrong. The special significance of this case is that it also violates the assumptions of the computer estimation procedure in such a manner that it produces exactly the same estimate as the human subjects. This is strong evidence that the algorithm people employ to estimate illuminant direction is similar to the one described above.

B. Using The Illuminant Direction To Type The Surface

Once we have an estimate of L for a region, we can use this estimate as a basis for

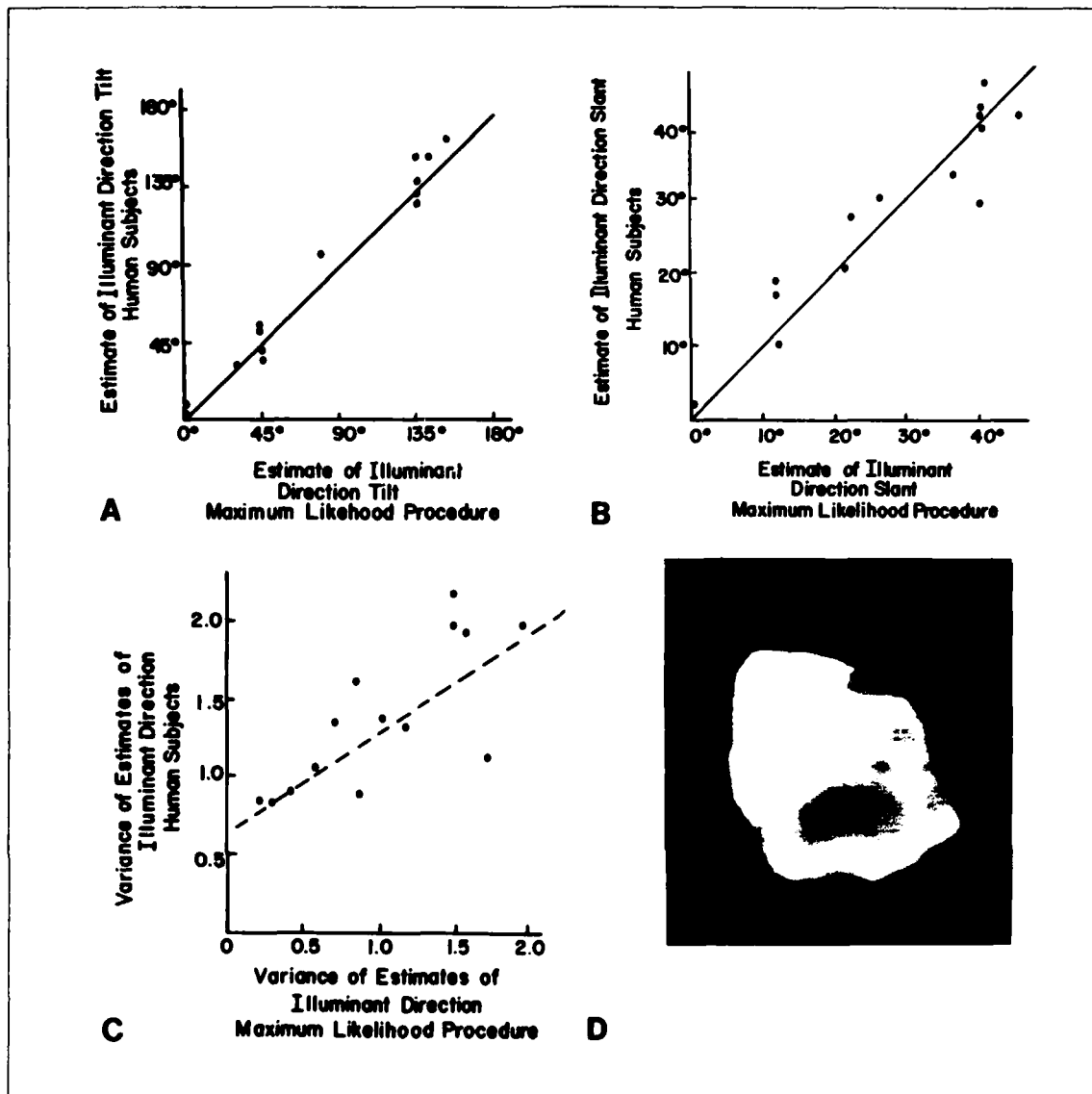


Figure 4. A comparison of human and computer estimates of illuminant direction in images of natural objects. Part (a) shows the comparison for the tilt component of the illuminant direction, and (b) shows the comparison for the slant component of the illuminant direction. Part (c) shows the relationship between the variance of human estimates of illuminant direction and the variance of the computer's estimate of illuminant direction. There is a correlation of 0.63 between the variances of the two sets of estimates, significant at the $p \leq 0.05$ level. The dashed line is the linear regression line relating the variance of the two estimation procedures. Part (d) is a picture of a rock for which both human estimates of illuminant direction and the maximum likelihood estimate agreed, but were objectively wrong. Actual illumination direction is top right, not top left as reported by all but two of the fifteen human subjects.

acquiring further information about the image formation process. One important use of L is to provide sufficient constraint to identify the surface as convex, concave or saddle — thus completing the typing of the surface.

Figure 5 contains an example of the "crater illusion." In this image, the shadow information is not prominent enough to determine the illuminant direction; consequently,



Figure 5. The Crater Illusion. Pictures of craters can look like bumps instead of depressions if we imagine the light source to be at the bottom of the picture rather than at the top; to see this, turn the figure upside down. This picture is of an ash cone in the Hawaiian Islands (courtesy of W. Richards).

no matter how the picture is turned the illumination is always (by default) seen to be coming from above¹³ Thus, the direction of illumination (relative to the image) changes as the image is turned upside down. When the direction of illumination changes, the convexity of the imaged surface also changes — thereby demonstrating that people use the direction of illumination to determine the convexity of the surface.

How can information about the direction of illumination be used to determine the surface convexity? When we invert Figure 5, the sign of dI as we move toward the apparent illuminant reverses, because the perceived image-plane component of the illuminant direction shifts by $\pi/2$ radians. Let us consider what the sign of dI tells us about the convexity of the imaged surface.

Equation (2) shows that the change in image intensity dI is dependent upon dN , the change in the surface normal:

$$dI = d(\rho\lambda(N \cdot L)) = \rho\lambda(dN \cdot L) + \rho\lambda(N \cdot dL) = \rho\lambda(dN \cdot L) \quad (2)$$

It turns out that dN is always perpendicular to N , which can be shown by observing that

$$2dN \cdot N = d(N \cdot N) = d(1) = 0$$

¹³This is an example of the above-mentioned $\pm\pi/2$ ambiguity in illuminant direction.

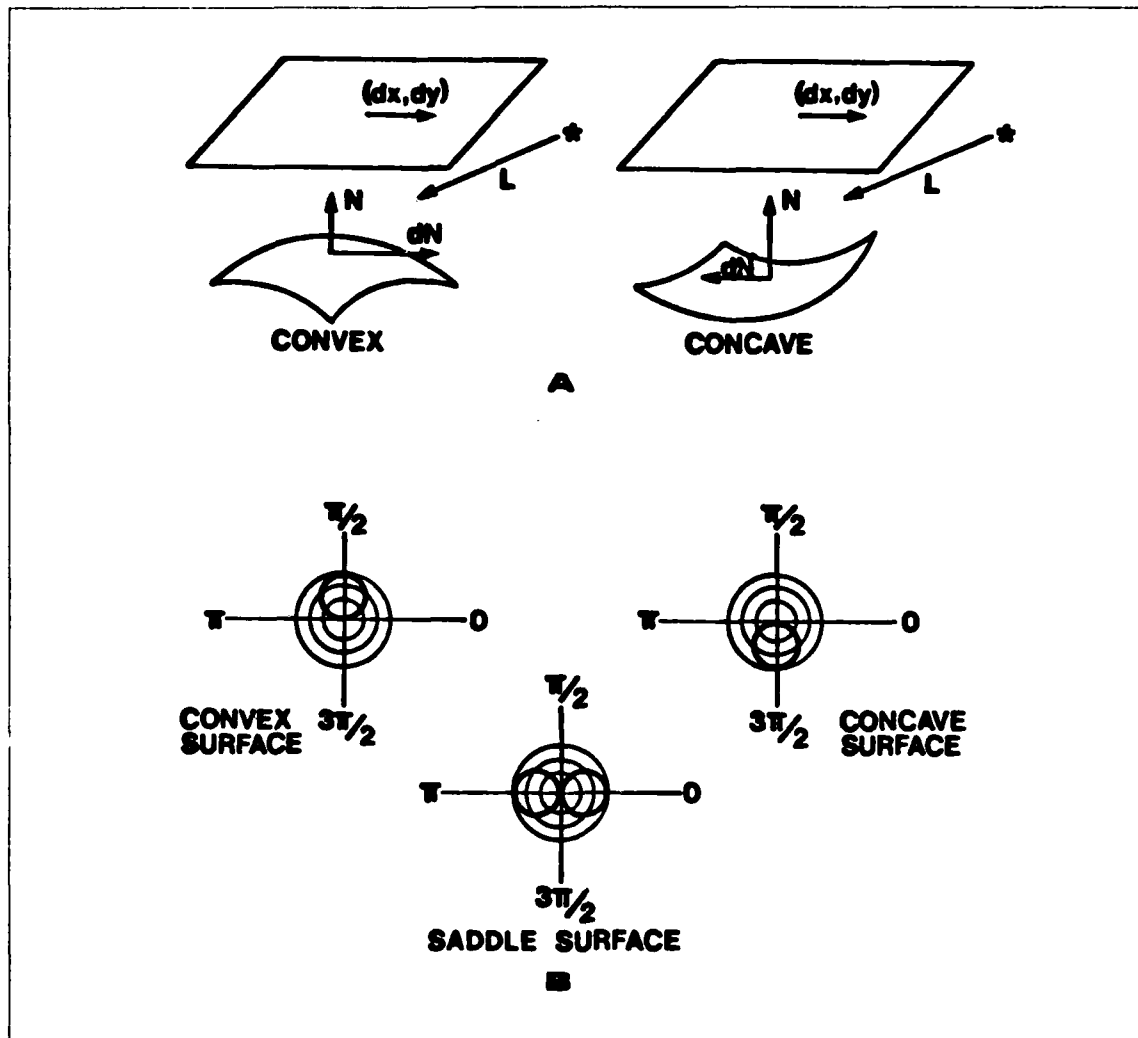


Figure 6. Estimation of surface type. (a) For a convex surface, dN measured along image direction (dx, dy) typically points in the direction $(dx, dy, 0)$, so that, if the image direction (dx, dy) is toward the light source, then $dI = \rho \lambda dN \cdot L$ is positive. For a concave surface, dN measured along image direction (dx, dy) typically points in the direction $(-dx, -dy, 0)$, so that $dI = \rho \lambda dN \cdot L$ is negative. Thus, the sign of dI in relation to the illuminant direction gives an estimate of the surface convexity along that direction. (b) The illuminant direction may be used to provide sufficient constraint to determine the qualitative type of surface. Each type of surface has a generic appearance, which may be characterized by the angle between τ_0 , the direction in which $dI = 0$, and τ_L , the illuminant direction. The distribution of $\tau_0 - \tau_L$ is shown for each surface type, assuming that the change in surface normal is isotropically distributed, and taking $dI > 0$ to the right of τ_0 . It can be seen that the appearance of the different types does not overlap much, so that a good identification of the surface type within the region may be made from this angle.

Whether dI is positive or negative along a particular direction depends upon whether dN points toward or away from the illuminant direction L . This is illustrated in Figure 6 (a).

If we assume that change in surface orientation is isotropically distributed within an image region, then, for a convex surface, dI measured in the image direction (dx, dy) will typically be positive if (dx, dy) is toward L . The sign of dI is positive because, for a convex

surface, dN measured along image direction (dx, dy) points on the average in the direction $(dx, dy, 0)$, so that $dI = \rho \lambda dN \cdot L$ is positive. In contrast, if the surface is concave dI will typically be negative because, for a concave surface, dN measured along image direction (dx, dy) points on the average in the direction $(-dx, -dy, 0)$, so that $dI = \rho \lambda dN \cdot L$ is negative. Thus, the sign of dI as we measure toward and away from the light source gives us an estimate of the convexity of the surface in that direction.

Unfortunately, dN measured along (dx, dy) does not usually point precisely along either the direction $(dx, dy, 0)$ or $(-dx, -dy, 0)$. Thus, even if we are given L and N , there remain too many unknown factors to establish the surface type with certainty. Each surface type, however, does have a typical or generic appearance. Therefore, given the tilt of L and the assumption that change in surface orientation is distributed isotropically within a region, we can estimate the surface type by observing the sign of dI measured toward and away from the illuminant.

Sufficient information for estimating the surface type is provided by the angle between τ_0 , the direction along which $dI = 0$, and τ_L , the tilt of the illuminant direction, as the sign of dI is positive on one side of τ_0 , negative on the other. Thus, knowing τ_0 and τ_L enables us to estimate the surface type. Figure 6 (B) shows the probability distribution of τ_0 for each surface type given τ_L , the tilt of the illuminant direction¹⁴, and the assumption that surface orientation is isotropically distributed within the region. As can be seen by comparing the overlap between these probability distributions, the likelihood of a correct identification is quite good.

Note that the ambiguity of $\pm\pi/2$ in the estimation of illuminant direction tilt leads to a global convexity/concavity ambiguity. Thus, just as with human perception, when a scene is sufficiently simple as to make L uncertain, the direction of illumination may be "switched" by $\pi/2$, which causes all the convexity/concavity determinations to change, as in Figure 5.

C. Estimation Of Surface Orientation

Although in real images relatively few points are umbilical and relatively few surfaces are Lambertian, the solution for surface tilt turns out to be fairly robust. The slant equation, however, depends critically on equal surface curvatures and on exact knowledge of the surface tilt. We must look further to find an estimator of surface slant that will be generally serviceable.

When a patch of surface is slanted away from the viewer, projection foreshortening occurs along the direction in which the surface tilts, causing an apparent increase in the surface curvature along that direction. This results in an increase in image curvature, i.e., the second derivative of the image intensities. Thus, for umbilical points (where the surface curvature is constant), the direction in which the second derivative of image intensity is greatest turns out to be the tilt of the surface. The slant of the surface can be measured by the amount of increase in image curvature.

The fact that increasing the surface slant results (all else being constant) in increased image curvature suggests that a measure of image curvature might be a good estimator of

¹⁴These distributions were determined by means of a Monte Carlo simulation.

slant. Image curvature, however, also depends on surface albedo, strength of illumination, surface curvature, and other factors. Still, if we investigate a homogeneous, uniformly lit region of a natural image we find that there is a good correlation between the values of a measure of image curvature such as the Laplacian ($\nabla^2 I$) and the surface slant. If the surface albedo or the illumination changes, however, there will be large changes in the Laplacian that have nothing to do with the surface slant — because the Laplacian values are directly dependent upon the surface albedo ρ and the illuminant strength λ .

If we divide the Laplacian values by the image intensity values, we can remove the dependence on ρ and λ , thus eliminating two of the most important confounding factors (see Equations (2) and (3)). The division of $\nabla^2 I$ by I also introduces a factor that is dependent upon the illuminant direction; however, this dependency does not seem to affect performance seriously — especially in natural imagery where there is a large amount of diffuse and reflected light. Thus, the division of $\nabla^2 I$ by I yields a measure that depends primarily upon the surface curvature and surface slant. Thus, we are led to the following estimator of surface slant, which is analyzed in the appendix.

Proposition 6. Given the image of an umbilical point on a Lambertian surface and R , the radius of surface curvature, the following is an estimate of z_N , the z component of the surface normal, equal to the arccosine of the surface slant:

$$z_N \approx -R^{-1} \left(\left| \frac{\nabla^2 I}{I} \right| - R^{-2} \right)^{-\frac{1}{2}}$$

This estimate of surface slant turns out to be much more robust than the umbilical-point solution for surface slant, degrading slowly as the surface curvatures become progressively more unequal or as the reflectance function becomes non-Lambertian.

Estimation of R . To use this estimator, the constant R must be determined. A good estimate of the mean R within an image region can be made by applying the constraint that the resulting z_N must satisfy the inequality $0 \geq z_N \geq -1$ — i.e., visible surfaces must be facing the viewer. We can determine a likely value for R by using this constraint and the equation for z_N in light of the range of values of $\nabla^2 I/I$ within a region.

We can then *assume* that the estimated value of R holds throughout the region, and thus obtain an estimate of intraregional slant. If the variance of R is small we will obtain a good estimate of surface orientation. It can happen, however, that the value of R will vary considerably from point to point — unless we can place bounds on the range of R so that its variance is reduced to an acceptable level.

Using the values of $d^2 I$ to identify planar and near-planar regions (as discussed in the previously), we can place a bound on the minimum value of R . We can also place a bound on the largest value of R by blurring image of the region in which slant is to be estimated¹⁵ Such blurring also has the effect of removing highlights, specularities, marks, textures, and the like, thus making the imaged surface more homogeneous and Lambertian. By bounding

¹⁵Noting that $I \otimes G = \rho \lambda (\mathbf{N} \cdot \mathbf{L}) \otimes G = \rho \lambda ((\mathbf{N} \otimes G) \cdot \mathbf{L}) = \rho \lambda (\tilde{\mathbf{N}} \cdot \mathbf{L})$ where G is a two-dimensional Gaussian and \otimes designates convolution, we see that a smoothed version of I may be considered the image of a surface with normal $\tilde{\mathbf{N}} = \mathbf{N} \otimes G = (N_x \otimes G, N_y \otimes G, N_z \otimes G)$, i.e., a smoothed version of the original surface.

R the variance can apparently be reduced to acceptable levels¹⁶, so that one may expect to obtain a useful estimate of surface shape within a homogeneous area.

Estimating tilt from the slant estimates. The umbilical-point solution for the surface tilt is the direction in which d^2I attains its maximum. To calculate the maximum, therefore, we require either the values of d^2I along many different orientations, or quite accurate values of¹⁷ I_{xx} , I_{yy} and I_{xy} . This is a fair number of image convolutions; besides, biological visual systems seem to manage with only $\nabla^2 I$ -like convolutions. It is, therefore, worth inquiring whether there exists a method of determining d^2I along a particular direction from the values of $\nabla^2 I$.

Let us examine the convolution filters required to calculate d^2I and $\nabla^2 I$, the differential quantities used to define the tilt and slant estimates, respectively. We can calculate the second derivative d^2I in the x direction by convolving the image with $d^2G(x, y, \sigma)/dx^2$, where $G(x, y, \sigma)$ is a two-dimensional Gaussian in the variables (x, y) with variance σ . Similarly, we can calculate the Laplacian $\nabla^2 I$ by convolving $\nabla^2 G(x, y, \sigma)$ with the image¹⁸. These two filters are closely related: if we sum $d^2G(x, y, \sigma)/dx^2$ and its 90° rotation, $d^2G(x, y, \sigma)/dy^2$, we obtain $\nabla^2 G(x, y, \sigma)$. We can obtain an approximation to the second-derivative filter $d^2G(x, y, \sigma)/dx^2$ by using a weighted sum of several Laplacians along a straight line in the perpendicular y direction, e.g.,

$$d^2G(x, y, \sigma)/dx^2 \approx \sum_{\epsilon} G(\epsilon, \sigma) \nabla^2 G(x_0, y_0 + \epsilon, \sigma)$$

where $G(\epsilon, \sigma)$ is a one-dimensional Gaussian, and $G(x_0, y_0, \sigma)$ designates a Gaussian centered about the point (x_0, y_0) . In this manner we can obtain a close approximation to $d^2G(x, y, \sigma)/dx^2$ from $\nabla^2 G(x, y)$ filters (see [15], [16]). Applying this result we see that if we were to sum the quantity $\nabla^2 I/I$ (the input data for the slant estimator) along a straight line, we would obtain an approximation to d^2I/I . This approximation allows us to compute the direction of maximum d^2I from the slant estimation data without additional convolutions; we need only find the orientation along which the sum of $\nabla^2 I/I$ is a maximum.

In practice this approximation to d^2I results in slightly better performance than using the filter that corresponds exactly to d^2I . The difference arises primarily in low-slant regions where the slant estimator (and thus its gradient and this approximation) is more stable than the straightforward tilt estimator.

Evaluation of the surface orientation estimate with an analytic model. To ascertain how well the slant and tilt estimators might be expected to perform under ideal

¹⁶Although we can reduce the variance of R , we cannot remove systematic bias. Thus, for example, if our viewpoint and the surface shape such that the surface curvature varies inversely with the surface slant (e.g., a parabolic solid viewed point-on), we will obtain a poor estimate of slant. It is worth noting that people also perform poorly under such conditions. Luckily, such arrangements are unusual in natural scenes because surface slant depends on viewpoint, unlike surface curvature; thus, the two are rarely inversely related.

¹⁷Possession of these three values allows analytic solution for the direction of maximum d^2I ; the solution is shown in the appendix.

¹⁸These convolutions may be regarded as calculating the exact values for a blurred version of the image; or, as mentioned earlier, the blurred image may be regarded as an exact image of a smoothed version of the original scene.

conditions, a computer program was written that used the analytic formulas to calculate the derivatives for images of a wide range of ellipsoidal solids. A Lambertian reflectance function was assumed and a wide range of illumination directions used. Three sets of solid shapes were utilized. The first set consisted only of spheres, so as to test the validity of the program. The second set consisted of ellipsoidal solids with a ratio of principal curvatures which ranged from 2:1 through 1:1 to 1:2. The third set consisted of ellipsoidal solids with a ratio of principal curvatures which ranged from 10:1 through 1:1 to 1:10. Thus, the third set of solids encompassed shapes ranging from almost completely cylindrical to spherical. Points were then sampled evenly from across the entire imaged surface and error statistics computed.

A summary of results for the tilt estimator over the three sets of solids, $\kappa_1 = \kappa_2$ (a sphere), $\kappa_1/\kappa_2 = 2:1$ (shapes between elongated eggs and spheres, i.e., common non-cylindrical shapes) and $\kappa_1/\kappa_2 = 10:1$ (shapes between cylinders and spheres, i.e., virtually all ratios of curvatures) is shown in Table 1. The direction of surface tilt was computed by using the approximation to the direction of maximum d^2I discussed in the appendix. All error figures are given in radians.

Table 1. Tilt estimator over all surface slants

Ratio Of Curvatures	Error: Bias, Variance	Correlation
$\kappa_1 = \kappa_2$	0.00, 0.052	0.891
$\kappa_1/\kappa_2 = 2:1$	0.00, 0.097	0.786
$\kappa_1/\kappa_2 = 10:1$	0.00, 0.114	0.742

Although the tilt estimator of Proposition 3 performs perfectly on spheres, the approximation used here shows some small errors. This loss of accuracy is offset by the greater stability that the approximation exhibits in low-slant regions. This table shows that as the range of curvatures increases the performance of the estimator degrades considerably. However, it is only in high-slant regions that errors in the tilt cause serious miscalculations in determining the surface shape; therefore, if the tilt estimator performs well in these regions the resulting shape estimate will still be accurate. Table 2 summarizes the tilt estimator's performance in the critical high-slant regions.

Table 2. Tilt estimator over surface slants greater than 30°

Ratio Of Curvatures	Error: Bias, Variance	Correlation
$\kappa_1 = \kappa_2$	0.00, 0.020	0.950
$\kappa_1/\kappa_2 = 2:1$	0.00, 0.066	0.835
$\kappa_1/\kappa_2 = 10:1$	0.00, 0.070	0.816

Table 2 describes the performance of the tilt estimator when the surface slant is greater than 30° . It can be seen that the tilt estimates remain quite reasonable for both the 2:1 and the 10:1 range of curvatures. Thus, the tilt estimator makes most of its errors in

low-slant regions, where such errors are relatively unimportant. Furthermore, if the slant estimator provides consistent estimates, we should be able to use it to distinguish between more reliable and less reliable tilt estimates.

Table 3. The slant estimator

Ratio Of Curvatures	Error: Bias, Variance	Correlation
$\kappa_1 = \kappa_2$	0.201, 0.012	0.924
$\kappa_1/\kappa_2 = 2:1$	0.182, 0.034	0.843
$\kappa_1/\kappa_2 = 10:1$	0.137, 0.101	0.674

Table 3 shows that the slant estimator, although biased, performs quite well on spheres. As the range of curvatures increases, the performance of the estimator degrades — but, even for the 10:1 range of curvatures, it is still good. For all these cases one estimate of R was used; therefore, in all except the case of spheres the estimated R is actually in error — in some instances by a factor of 10. This seems to indicate that the slant estimator is remarkably robust.

Table 4. The unbiased slant estimator

Ratio Of Curvatures	Error: Bias, Variance	Correlation
$\kappa_1 = \kappa_2$	0.00, 0.004	0.967
$\kappa_1/\kappa_2 = 2:1$	0.00, 0.006	0.948
$\kappa_1/\kappa_2 = 10:1$	0.00, 0.028	0.796

The bias of the slant estimator that appears in Table 3 also shows up in the equations discussed in the appendix; where it is also explained how the bias may be removed. When the slant estimator is made unbiased, its accuracy becomes even better, as shown in Table 4. It turns out that the performance of the slant estimator is approximately as good for regions of low slant as for those of higher slant. Therefore, the slant estimate produced by this estimator is useful in assessing the tilt estimator's reliability.

Evaluation on natural images. The surface orientation estimator (the "shape algorithm") has been tested on several natural images, and four such examples will be presented here. The shape algorithm produces estimates of the surface orientation; it was found, however, that displays of the estimated surface orientation do not allow an observer to evaluate the performance of the algorithm adequately. Therefore, for purposes of exhibiting the performance of the algorithm, the shape algorithm's estimates of surface orientation were integrated to produce a relief map of the surface. As these relief maps were found to give observers an adequate impression of the estimated surface shape, they constitute the output shown for the examples presented in this paper even though integration is not part of the shape algorithm *per se*.

Figure 7 (a) shows the image of a log, together with the relief map generated from the shape algorithm's estimates of surface orientation. Figure 7 (b) shows the image of a

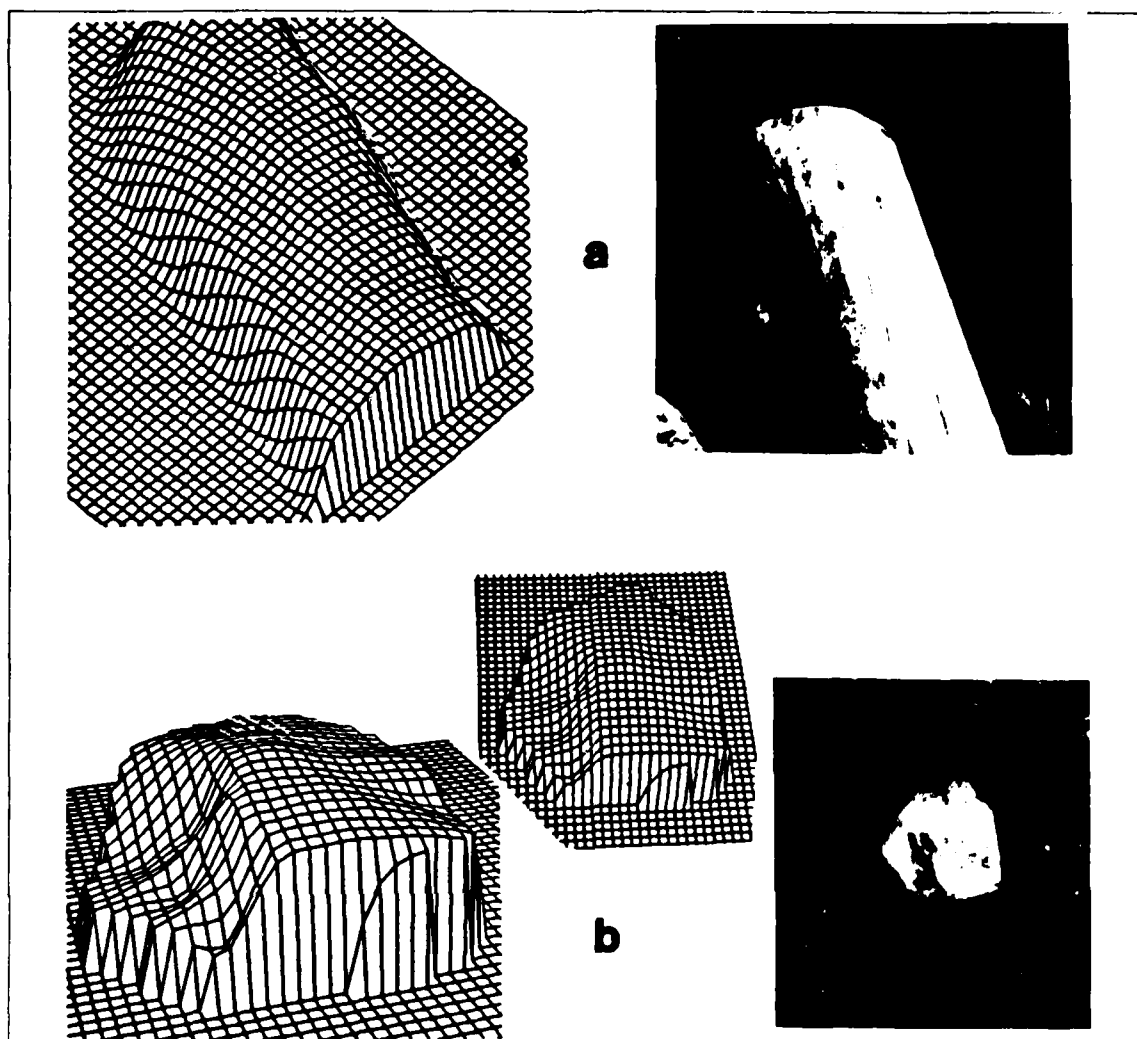


Figure 7. Evaluation on two natural images. (a) An image of a log and the relief map generated by the shape algorithm for that image. (b) An image of a rock and the relief map generated by the shape algorithm for that image.

rock, together with the relief map generated from the shape algorithm's estimates of surface orientation. The relief maps in Figure 7 (a) and 7 (b) correspond closely to the actual shapes of these two objects. The reader should compare his impression of shape from the images with the relief maps of Figure 7 (a) and 7 (b).

Figure 8 shows (a) the digitized picture of a small portion of a face (belonging to a woman named Lisa), and (b) a relief map of the surface slant estimated for that image (eye and eyebrow regions were masked out by hand). No relief map of the estimated surface shape is shown because the complexity of the shape made it difficult to integrate the slant and tilt estimates. In this slant-map representation regions with higher relief face toward the viewer, while lower relief regions face away from him. Note that many important details of the surface shape are apparent in this representation; for instance, the structure of the nose, the cheeks and the eyebrow ridges is plainly visible.

Figure 9 shows (a) the digitized image of Tuckerman's ravine (a skiing region on Mt.

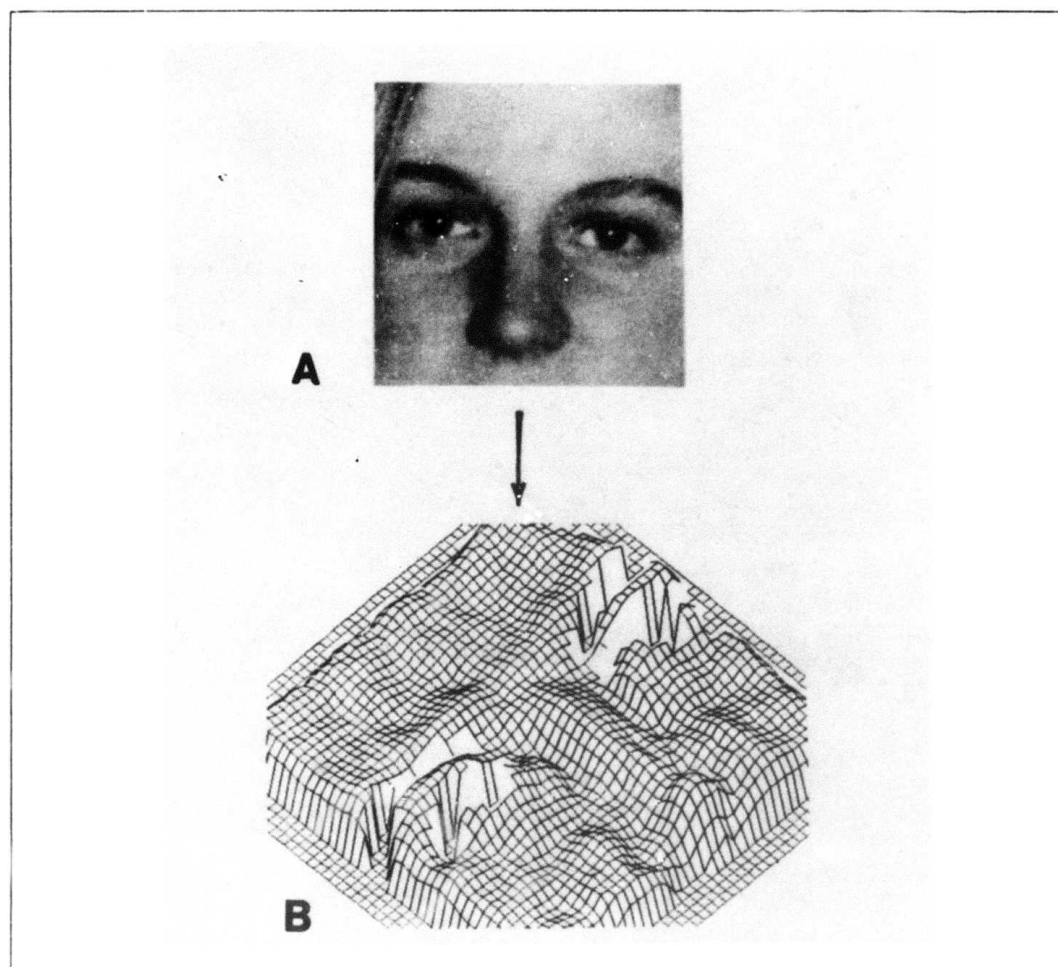


Figure 8. The Lisa Image. (a) The digitized image of Lisa and (b) a relief map showing the estimated surface slant in this image. High-relief areas in this representation correspond to regions facing the viewer, low-relief regions to regions slanting away from the viewer. Eye regions and eyebrow regions were masked out by hand. The integration of estimated slant and estimated tilt to show surface shape proved difficult because of the complexity of the surface. Note that many important details of the surface shape are apparent in this representation; for instance, the structure of the nose, cheeks and eyebrow ridges is plainly visible.

Washington, in New Hampshire), (b) a relief map showing a side view of the estimated surface shape, obtained by integrating the slant and tilt estimates. This relief map may be compared directly with a topographic map of the area; when we compare the estimated and actual shape, we find that the roll-off at the top of Figure 9 (b) and the steepness of the estimated surface are correct for this surface¹⁰; this area of the ravine has a slope that averages 60°.

The comparison with a topographic map also shows that the relief of the lower right-hand portion of the image is somewhat underestimated. When people are asked to look at this image, however, it becomes clear that they also fail to perceive the shape of the ravine

¹⁰And people ski down this incline!

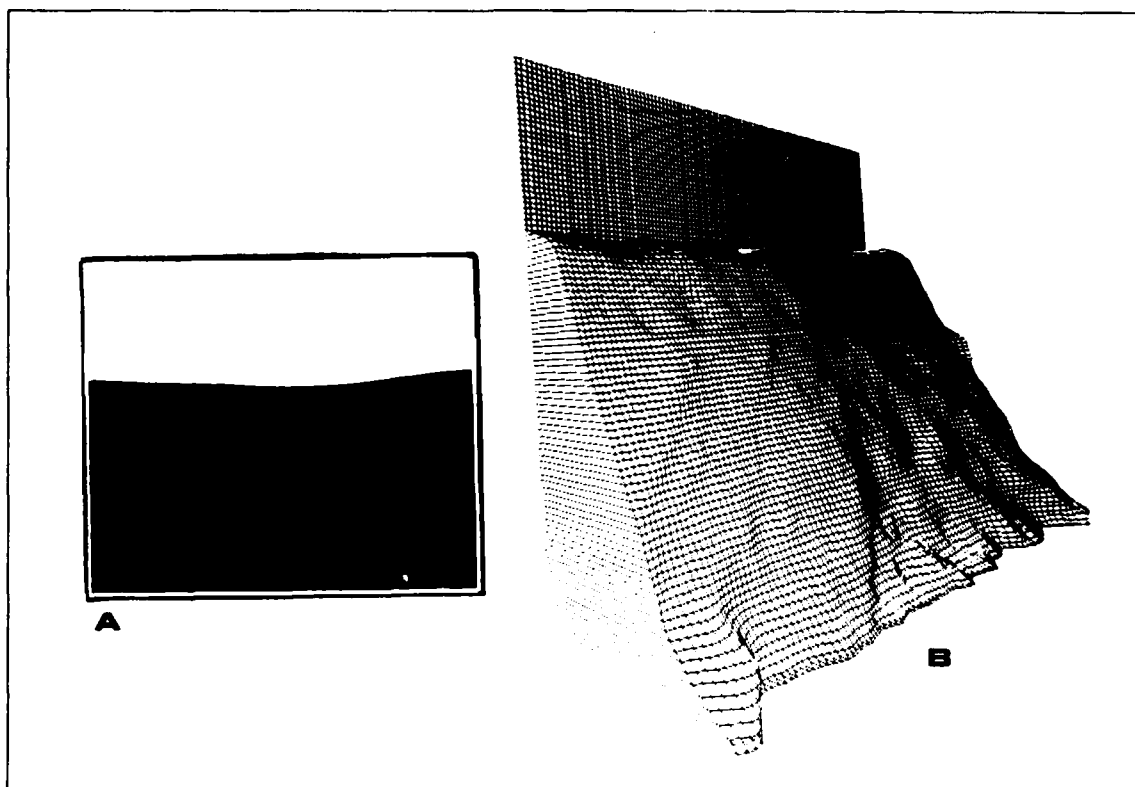


Figure 9. Tuckerman's Ravine. (a) The digitized image of Tuckerman's ravine and (b) a relief map showing a side view of the surface estimated for this image. Comparing a topographic map of the area with the estimated surface shape, we find that the roll-off at the top of (b) and the steepness of the estimated surface are correct for this surface. However, the relief of the lower right-hand portion of the image is somewhat underestimated. The underestimation of relief is similar to human perception of this image.

correctly: they also underestimate the relief of the lower right portion of this image²⁰.

Evaluation On An Electron Microscope image. In addition to natural images, the electron microscope (EM) image shown in Figure 10 (a) was selected from the book *Magnifications* by D. Scharf [6]. People can use the shading information in EM images to perceive shape, as Figure 10 (a) confirms. This is surprising because these images have a reflectance function not found in natural scenes. This image, therefore, provides a critical test of the similarity between the human use of shading and this estimator of surface orientation.

Ikeuchi and Horn [3] measured the reflectance function for this image and found that the image intensities may be reasonably well described by

$$I = k(N \cdot V)^{-1}$$

where k is approximately 0.8. If we carry out the required computations, we see that the tilt is still the direction along which d^2I is greatest, and that the z -component of the surface normal is approximately proportional to $\nabla^2 I / I$, as in normal images. We can thus expect to

²⁰However, when people are able to view the original higher-resolution image or the entire image they perceive the surface correctly.

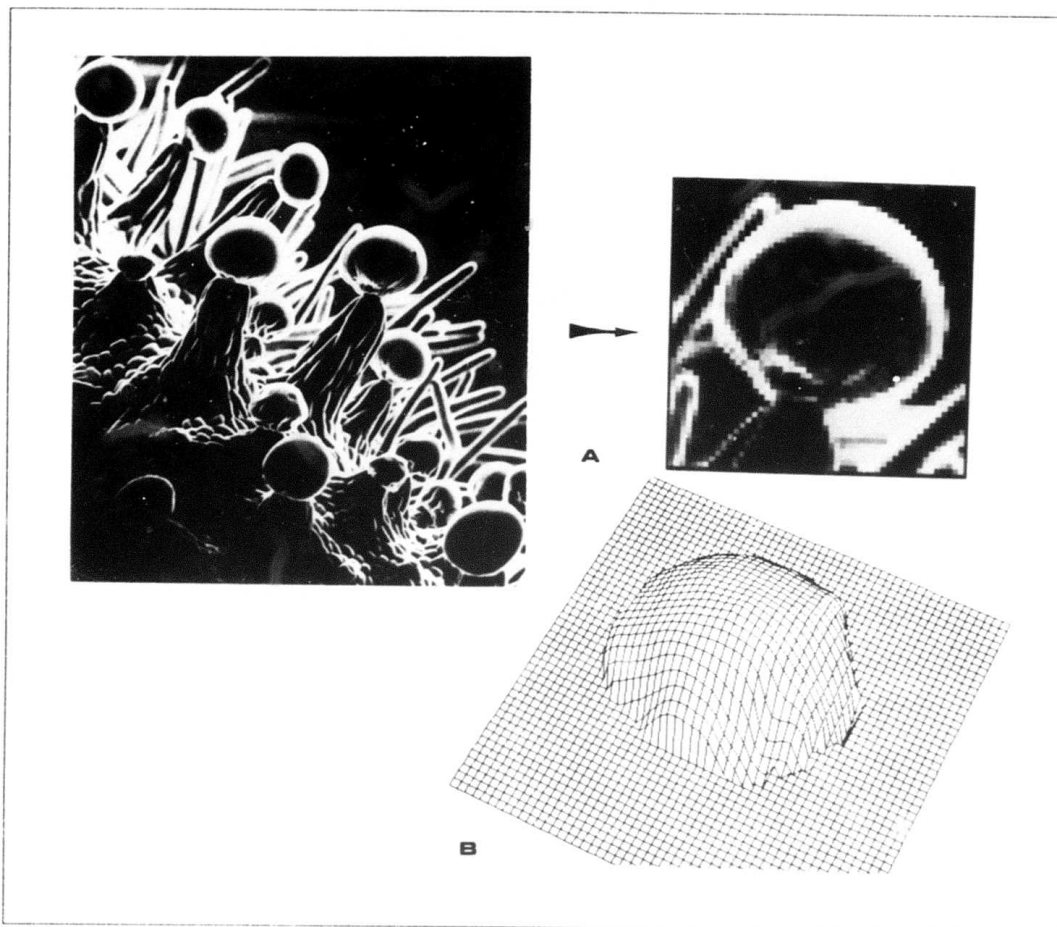


Figure 10. An electron microscope image. Part (a) is an electron microscope image of resin nodules on a flower of *cannabis sativa*, and the portion of the image for which shape was estimated. Part (b) shows a relief map showing a side view of the surface estimated for this image. The fact that both people and this algorithm can correctly use the shading information in electron microscope images may have important implications for understanding human vision.

obtain a reasonable shape estimate for EM images by using the same estimation technique developed for normal images.

Figure 10 (b) the portion of 10 (a) for which shape was estimated. Figure 10 (c) is a relief map showing a side view of the estimated surface shape, again obtained by integrating the slant and tilt estimates for this image. It can be seen that the estimated surface shape is quite accurate.

IV. Discussion

The preceding portions of this paper have shown that it is possible to obtain useful estimates of scene properties from natural images by using a local analysis of image shading. The analysis does *not* assume that any scene information is known beforehand; no knowledge

of scene characteristics or boundary conditions is used. Thus, the techniques are applicable to "raw" images. Because no *a priori* information is assumed, however, the recovery of information about the scene is necessarily imprecise.

One major problem which is inherent in obtaining shape from shading under general viewing conditions is that the measured image intensities are not equal to the image irradiance. Film, video camera, and other image transcription methods produce image intensity measurements which are (generally) non-linear transformations of the image irradiance. It is, therefore, surprising to note that humans have no problem in utilizing the shading in such transformed images even though the relationship between the transformed image and the original image irradiance is unknown. Thus, any shape-from-shading technique which will be as generally useful as the human capacity must function *despite* such transformations of the data.

The shape-from-shading techniques described in this paper are relatively unaffected by smooth, monotonic transformations of the image data — in marked contrast to previous methods of inferring shape from shading. This robustness is achieved by dividing the Laplacian of the image intensities by the intensities themselves, thus removing the primary effects of any multiplicative terms in the image irradiance equation. The division also removes the effects of any linear scaling of the image intensity. Thus, division of the Laplacian by the intensity compensates for any transformation of the image irradiance which is locally approximately linear.

What is the use of a reasonably accurate, but certainly not infallible, local estimate of scene properties? Several potential applications spring to mind: to provide an initial "guess" for a more global shading analysis [20], to constrain stereo matching by providing a qualitative estimate of shape, or to help in the estimation of albedo. One other use that I have begun examining is classification of the type of imaged contours. This serves as a good demonstration of the potential usefulness of a reasonably accurate local estimate of surface orientation.

Once we are given the location of a contour we should be able to use our local estimate of surface orientation to ascertain whether a contour is a smooth occluding contour (i.e., a contour formed by the surface curving smoothly out of sight, such as is found at the edge of an image of a sphere) by checking whether our estimates of surface orientation are appropriate for that contour. If the surface adjoining one side of a contour has a large slant and a tilt perpendicular to the contour, then it is likely a smooth occluding contour. On the other hand, if the estimated slant is small, or if the tilt is not perpendicular, then it is quite probably not a smooth occluding contour.

Figure 11 shows the results of applying this testing strategy to the contours extracted from two natural images. Part (a) of this figure shows the Moore sculpture image, and part (d) the Tuckerman's ravine image. Parts (b) and (e) depict the discontinuity contours found in these images. Parts (c) and (f) of this figure show the contours that were adjoined by regions whose estimated surface orientation was consistent with the contours' being a smooth occluding contour. When we compare the contours identified as smoothly occluding with the original images, we find that this criterion is quite apt in identifying the smooth occluding contours in these images.

Biological Vision Systems. The shape-from-local-shading theory presented in this paper

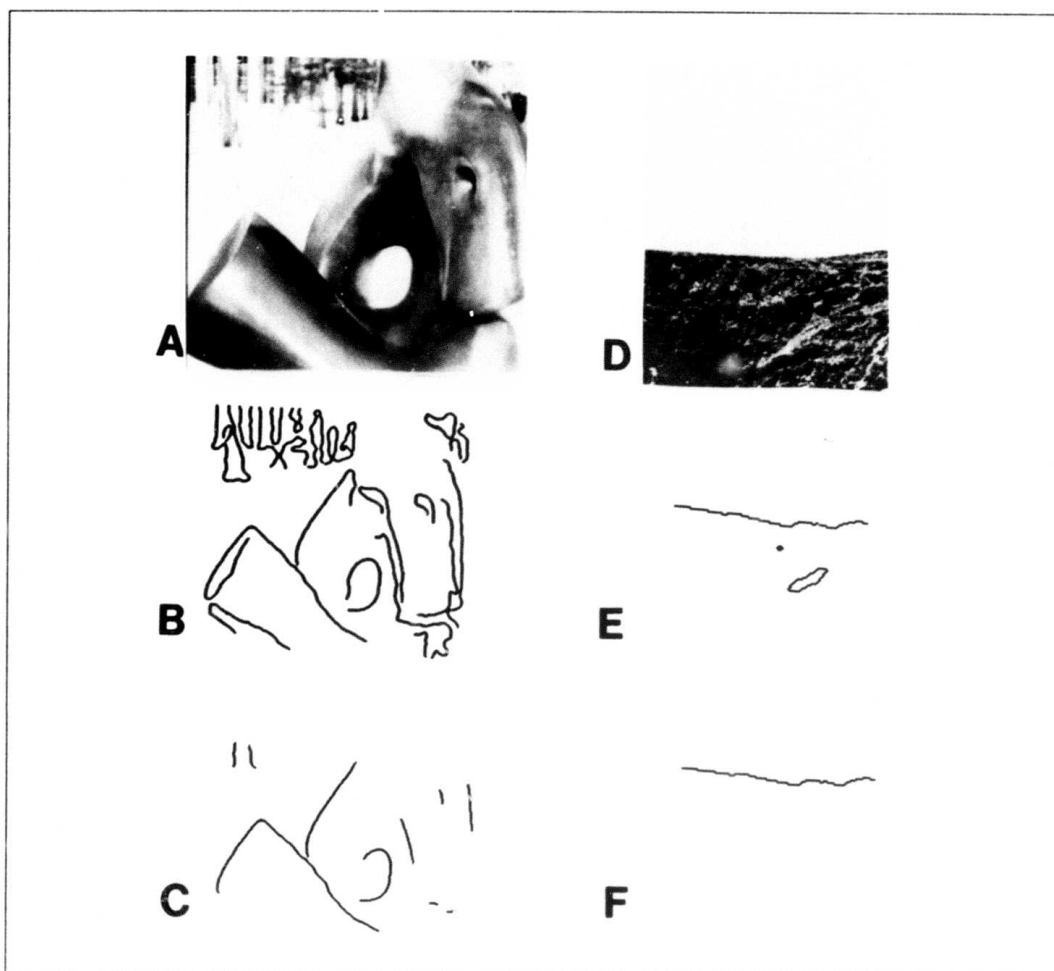


Figure 11. Results of typing the contours extracted from two natural images. Part (a) of this figure shows the Moore sculpture image, and part (d) shows the Tuckerman's ravine image. Parts (b) and (e) show the discontinuity contours found in these images. Parts (c) and (f) show the contours which were adjoined by regions whose estimated surface orientation was consistent with the contour being a smooth occluding contour. When we compare these contours to the original images, we find that this criterion does quite well at identifying the smooth occluding contours in these images.

seems to have considerable utility as a model of one aspect of the functioning of biological visual systems. It is known from studies of neurophysiology [8], [9] and human psychophysics [10], [11] that the retinal receptive fields in mammals have a center-surround organization that is well modeled by the filter $\nabla^2 G(x, y, \sigma)$. In addition, the responses of these retinal neurons are logarithmically scaled by the intensity, so that their response r to an image point $I(x_0, y_0)$ can reasonably be modeled by the following convolution and division:

$$r = \frac{\nabla^2 G(x, y, \sigma) \otimes I(x_0, y_0)}{I(x_0, y_0)} \quad (5)$$

The quantity r is the measurement needed by the slant estimator. Thus, it seems that the information required to estimate slant is present in the output of the mammalian retina.

It is also well established [12], [13], [14] that many (perhaps most) cortical neurons in the primary visual cortex have oriented receptive fields whose responses characteristics are closely modeled by the filter $d^2G(x, y, \sigma)/dx^2$. Moreover, there is very strong evidence [15], [16] that these cortical neurons are constructed by summing the center-surround receptive fields described by Equation (5), just as was done here. Thus, not only does it seem that the image data required by the tilt estimator is present in the mammalian primary visual cortex, but the information is apparently derived from the image by means of the same steps that have been employed here.

Acknowledgments. I would like to thank my fellow graduate students, including Dr. Andy Witkin, Donald Hoffman, Joseph Scheuhammer and Dr. Eric Grimson, who helped in the refinement and development of this work.

V. Appendix

In this section the previous five propositions will be proved. The strategy of proof will be to start with the six values for I , I_x , I_y , I_{xx} , I_{yy} and I_{xy} and then solve for surface orientation, curvature, illuminant direction and albedo times illuminant intensity under the assumption that the point is an umbilical point on a Lambertian surface. This solution will, at the same time, prove Propositions 1, 2, and 4. Some additional calculations using the results of this solution will then prove Propositions 3, 5 and 6.

A. Solution For Umbilical Points

Consider the surface of a sphere of radius R :

$$Z(x, y) = \sqrt{R^2 - x^2 - y^2}$$

This equation, with $R > 0$, $R \geq x \geq -R$ and $R \geq y \geq -R$, describes the set of all umbilical points. From this equation we see that $Z_x = -xT^{-1/2}$ and $Z_y = -yT^{-1/2}$ where $T = R^2 - x^2 - y^2$. Assume that the illuminant is unknown, so that we must consider all illumination directions $\mathbf{L} = (l_1, l_2, l_3)$. Then, if the surface is Lambertian, we have

$$I(x, y) = \rho \lambda \mathbf{N} \cdot \mathbf{L} = \frac{\rho \lambda (Z_x, Z_y, -1) \cdot (l_1, l_2, l_3)}{\sqrt{Z_x^2 + Z_y^2 + 1}} = \frac{\rho \lambda}{R} (-xl_1 - yl_2 - l_3 T^{1/2}) \quad (1)$$

where ρ is the surface albedo and λ is the illuminant intensity at the surface. The first and second derivatives are then

$$I_x = \frac{\rho \lambda}{R} (-l_1 + xl_3 T^{-1/2}) \quad (2)$$

$$I_y = \frac{\rho \lambda}{R} (-l_2 + yl_3 T^{-1/2}) \quad (3)$$

$$I_{xx} = \frac{\rho \lambda}{R} (l_3 T^{-1/2} + x^2 l_3 T^{-3/2}) \quad (4)$$

$$I_{yy} = \frac{\rho \lambda}{R} (l_3 T^{-1/2} + y^2 l_3 T^{-3/2}) \quad (5)$$

$$I_{xy} = \frac{\rho \lambda}{R} (xy l_3 T^{-3/2}) \quad (6)$$

Assume that the values of I , I_x , I_y , I_{xx} , I_{yy} and I_{xy} are known; we may now solve for surface orientation, curvature, illuminant direction and albedo times illuminant intensity.

Solution

Using Equation (6) to solve for $\frac{\rho \lambda}{R}$ we obtain

$$\frac{\rho \lambda}{R} = \frac{I_{xy} T^{3/2}}{xy l_3} \quad (7)$$

Using Equations (7) and (4) we obtain

$$I_{xx} = (l_3 T^{-1/2} + x^2 l_3 T^{-3/2}) \frac{\rho \lambda}{R} = (l_3 T^{-1/2} + x^2 l_3 T^{-3/2}) \frac{I_{xy} T^{3/2}}{xy l_3} = \left(\frac{T + x^2}{xy} \right) I_{xy} \quad (8)$$

and using Equation (7) with Equation (5) we obtain

$$I_{yy} = (l_3 T^{-1/2} + y^2 l_3 T^{-3/2}) \frac{\rho \lambda}{R} = (l_3 T^{-1/2} + y^2 l_3 T^{-3/2}) \frac{I_{xy} T^{3/2}}{xy l_3} = \left(\frac{T + y^2}{xy} \right) I_{xy} \quad (9)$$

Using Equations (8) and (9) we see that

$$T = \frac{xy I_{xx}}{I_{xy}} - x^2 = \frac{xy I_{yy}}{I_{xy}} - y^2 \quad (10)$$

Using Equations (9) and (10) we see that

$$\begin{aligned} I_{yy} &= \left(\frac{T + y^2}{xy} \right) I_{xy} \\ &= \left(\frac{\frac{xy I_{xx}}{I_{xy}} - x^2 + y^2}{xy} \right) I_{xy} \\ &= I_{xx} - \frac{x}{y} I_{xy} + \frac{y}{x} I_{xy} \end{aligned} \quad (11)$$

Letting $k = \frac{y}{x}$ we see that Equation (11) is a quadratic in k , which we can solve to obtain

$$k = \frac{-(I_{xx} - I_{yy}) \pm \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}}{2I_{xy}} \quad (12)$$

From this we may obtain the surface tilt $\tau = \tan^{-1} k$. Thus the tilt of the sphere's surface may be determined without knowledge of the illuminant direction, the illuminant strength, the surface albedo or the surface curvature. To prove Proposition 3, which stated that the tilt of the surface is in the direction of maximum $d^2 I$, it remains only to show that this solution for the surface tilt is the direction of maximum $d^2 I$. The remainder of the proof will be presented in the following subsection.

Is this solution unique? Equation (12) yields two possible solutions:

$$k_1 = \frac{y_1}{x_1} \quad k_2 = \frac{y_2}{x_2}$$

Note that we may also solve Equation (12) for k^{-1} ; this yields

$$k^{-1} = \frac{-(I_{xx} - I_{yy}) \pm \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}}{-2I_{xy}} \quad (13)$$

Equation (13) also gives two solutions, k_3 and k_4 ,

$$k_3 = k_1^{-1} = \frac{x_1}{y_1} \quad k_4 = k_2^{-1} = \frac{x_2}{y_2}$$

As the left-hand side of Equation (12) is the negative of the left-hand side of Equation (13) we find that either $k_3 = -k_1^{-1}$, which leads to a contradiction, or $k_3 = -k_2^{-1}$. Thus

$$k_1 = \frac{y_1}{x_1} = -k_2^{-1} = -\frac{x_2}{y_2}$$

This proves that the two solutions of Equation (12) are perpendicular. We shall see later that this allows us to discard one of the solutions, because it results in an illuminant direction that is behind the observed object.

Once the tilt r is known, the only remaining component of surface orientation is σ , the slant of the surface, which is equal to the arccosine of the z component of the surface normal, z_N . Noting that

$$z_N = \frac{-1}{\sqrt{Z_x^2 + Z_y^2 + 1}} = \frac{-\sqrt{R^2 - x^2 - y^2}}{R}$$

we find that we can use Equation (10) to determine the surface slant. We may average the two expressions for T in Equation (10) to obtain

$$T = R^2 - x^2 - y^2 = \frac{xy(I_{xx} + I_{yy})}{2I_{xy}} - \frac{(x^2 + y^2)}{2} = \frac{xy\nabla^2 I}{2I_{xy}} - \frac{(x^2 + y^2)}{2} \quad (14)$$

If we add $x^2 + y^2$ to both sides of Equation (14) we then obtain

$$R^2 = \frac{xy\nabla^2 I}{2I_{xy}} + \frac{(x^2 + y^2)}{2} \quad (15)$$

Thus, from Equations (14) and (15):

$$\begin{aligned} z_N^2 &= \frac{R^2 - x^2 - y^2}{R^2} \\ &= \frac{\frac{xy\nabla^2 I}{2I_{xy}} - \frac{(x^2 + y^2)}{2}}{\frac{xy\nabla^2 I}{2I_{xy}} + \frac{(x^2 + y^2)}{2}} \\ &= \frac{xy\nabla^2 I - (x^2 + y^2)I_{xy}}{xy\nabla^2 I + (x^2 + y^2)I_{xy}} \\ &= \frac{k\nabla^2 I - (k^2 + 1)I_{xy}}{k\nabla^2 I + (k^2 + 1)I_{xy}} \end{aligned} \quad (16)$$

From z_N^2 we can obtain the surface slant $\sigma = \cos^{-1} \sqrt{z_N^2}$. This concludes the proof of Proposition 4. Note that there is only one solution for the surface slant, as $0 \geq z_N \geq -1$. With both the slant and tilt determined, we now know the surface orientation. Once again, this has been accomplished without prior knowledge of illuminant direction, illuminant strength, surface curvature, or surface albedo.

We may now proceed to solve for the remaining unknowns. Because we know the surface orientation, we can compute $\chi = x/R$ and $\gamma = y/R$.

$$\chi = \cos r \sin \sigma \quad \gamma = \sin r \sin \sigma$$

The quantities χ and γ may be thought of as the x and y coordinates normalized to the unit sphere. Using χ and γ , we may define Γ , a unit sphere analogue to T

$$\Gamma = 1 - \chi^2 - \gamma^2 = \frac{T}{R^2}$$

Going back to Equation (7) we now know all of the terms on the left-hand side except l_3 , thus, let us write

$$\frac{\rho\lambda l_3}{R} = \frac{I_{zy}T^{3/2}}{xy} = \frac{I_{zy}\Gamma^{3/2}R}{\chi\gamma}$$

and so

$$\rho\lambda l_3 = \frac{I_{zy}\Gamma^{3/2}R^2}{\chi\gamma} \quad (17)$$

We may then substitute this equation into Equations (2) and (3) to give

$$\rho\lambda l_1 = \frac{I_{zy}\Gamma R^2}{\gamma} - I_x R \quad (18)$$

and

$$\rho\lambda l_2 = \frac{I_{zy}\Gamma R^2}{\chi} - I_y R \quad (19)$$

If we now convert Equation (1) to the variables χ , γ and Γ , we may substitute Equations (17), (18) and (19) into (1) to obtain a quadratic in R :

$$\begin{aligned} I &= \frac{\rho\lambda}{R}(-xl_1 - yl_2 - T^{1/2}l_3) \\ &= \rho\lambda(-\chi l_1 - \gamma l_2 - \Gamma^{1/2}l_3) \\ &= -\chi\left(\frac{I_{zy}\Gamma R^2}{\gamma} - I_x R\right) - \gamma\left(\frac{I_{zy}\Gamma R^2}{\chi} - I_y R\right) - \frac{I_{zy}\Gamma^{3/2}R^2}{\chi\gamma} \\ &= (\chi I_x + \gamma I_y)R - \frac{I_{zy}\Gamma R^2}{\chi\gamma} \end{aligned} \quad (20)$$

We may solve this for R to obtain

$$R = \frac{\chi\gamma\left((\chi I_x + \gamma I_y) \pm \sqrt{(\chi I_x + \gamma I_y)^2 - \frac{4I_{zy}\Gamma R^2}{\chi\gamma}}\right)}{2I_{zy}\Gamma R^2} \quad (21)$$

Because \mathbf{L} is a unit vector, we may now use Equations (17), (18) and (19) to determine $(\rho\lambda)^2$:

$$\begin{aligned} (\rho\lambda)^2 &= (\rho\lambda l_1)^2 + (\rho\lambda l_2)^2 + (\rho\lambda l_3)^2 \\ &= \left(\frac{I_{zy}\Gamma R^3}{\gamma} - I_x R\right)^2 + \left(\frac{I_{zy}\Gamma R^3}{\chi} - I_y R\right)^2 + \left(\frac{I_{zy}\Gamma^{3/2}R^2}{\chi\gamma}\right)^2 \end{aligned} \quad (22)$$

As $\rho\lambda \geq 0$, we may discard the negative root of $\sqrt{\rho\lambda}$, so that $\rho\lambda$ is uniquely determined. Using this value of $\rho\lambda$, we may now substitute into Equations (17), (18) and (19) to obtain l_1 , l_2 and l_3 .

Note that the signs of l_1 , l_2 and l_3 depend on the signs of χ and γ . Thus, in solving Equation (12) one of the two solutions will make $\chi\gamma$ negative making l_3 negative — which corresponds to an illuminant behind the observed surface. Therefore, only one of the two solutions of (12) is physically possible.

Because Equation (12) gives only the solution for y/x , there are two possible pairs (x_1, y_1) , (x_2, y_2) such that $k = y_1/x_1 = y_2/x_2$. These two solutions are each other's negative, i.e., $x_1 = -x_2$, $y_1 = -y_2$; consequently, the surface tilts for these two solutions are 180° apart so that one corresponds to a convex surface, the other to a concave surface. Because it is the signs of χ and γ that determine the signs of l_1 and l_2 , choosing one of the (x, y) pairs results in an illuminant direction that is overhead (i.e., $L \cdot (0, 1, 0) \geq 0$), while picking the other results in an illuminant direction that is below the viewing line. Thus, if we specify an illuminant direction which must be overhead and in front of the illuminated object, there is only one possible solution to Equation (12). The symmetry between the signs of x , y and l_1 , l_2 is commonly familiar as the crater illusion, in which the convexity of the surface changes as the perceived illuminant direction shifts from overhead to below the viewing line.

We have now solved for each of the unknown quantities, and, by so doing have shown that to each set of measurements I , I_x , I_y , I_{xx} , I_{yy} and I_{xy} there corresponds exactly one combination of surface orientation, curvature, (overhead) illuminant direction, and factor $\rho\lambda$ for a Lambertian umbilical point. This concludes the proofs of Propositions 1 and 2.

B. Proposition 3

In the preceding subsection it was shown that

$$r = \tan^{-1} k = \tan^{-1} \frac{y}{x}$$

where

$$k = \frac{-(I_{xx} - I_{yy}) \pm \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}}{2I_{xy}} \quad (12)$$

It remains to show that this solution is equivalent to the proposition that the tilt is the image direction in which d^2I is greatest.

We know that, given I_{xx} , I_{yy} and I_{xy} we may obtain these quantities in any other image plane coordinate system (x^*, y^*) that is a rotation of (x, y) by the angle ξ . First we note that

$$I_{x^*} = I_x \frac{dx}{dx^*} + I_y \frac{dy}{dx^*} \quad I_{y^*} = I_x \frac{dx}{dy^*} + I_y \frac{dy}{dy^*}$$

The standard rotation transformation is

$$\begin{aligned} x^* &= xc_\xi + ys_\xi \\ y^* &= -xs_\xi + yc_\xi \end{aligned}$$

where s_ξ and c_ξ are the sine and cosine of the angle ξ . The inverse of this rotation transformation is

$$\begin{aligned} x &= x^*c_\xi - y^*s_\xi \\ y &= x^*s_\xi + y^*c_\xi \end{aligned}$$

Thus,

$$\frac{dx}{dx^*} = c_\xi \quad \frac{dy}{dx^*} = s_\xi \quad \frac{dx}{dy^*} = -s_\xi \quad \frac{dy}{dy^*} = c_\xi$$

and so

$$\begin{aligned} I_x^* &= I_x c_\xi + I_y s_\xi \\ I_y^* &= -I_x s_\xi + I_y c_\xi \end{aligned}$$

Similarly,

$$\begin{aligned} I_{x^*x^*} &= \frac{d(I_x^*)}{dx} \frac{dx}{dx^*} + \frac{d(I_x^*)}{dy} \frac{dy}{dx^*} \\ I_{y^*y^*} &= \frac{d(I_y^*)}{dx} \frac{dx}{dy^*} + \frac{d(I_y^*)}{dy} \frac{dy}{dy^*} \\ I_{x^*y^*} &= \frac{d(I_x^*)}{dx} \frac{dx}{dy^*} + \frac{d(I_x^*)}{dy} \frac{dy}{dy^*} \end{aligned}$$

resulting in

$$\begin{aligned} I_{x^*x^*} &= I_{xx} c_\xi^2 + I_{yy} s_\xi^2 + 2I_{xy} s_\xi c_\xi \\ I_{y^*y^*} &= I_{xx} s_\xi^2 + I_{yy} c_\xi^2 - 2I_{xy} s_\xi c_\xi \\ I_{x^*y^*} &= -I_{xx} s_\xi c_\xi + I_{yy} s_\xi c_\xi + I_{xy}(c_\xi^2 - s_\xi^2) \end{aligned}$$

To find the direction for which d^2I attains its maximum, we find the angle ξ for which $I_{x^*x^*}$ attains a maximum over all rotations of the image plane coordinate system. As $I_{x^*x^*}$ is equal to

$$I_{x^*x^*} = I_{xx} c_\xi^2 + I_{yy} s_\xi^2 + 2I_{xy} s_\xi c_\xi$$

the maximum of $I_{x^*x^*}$ occurs at

$$\begin{aligned} 0 &= \frac{d(I_{x^*x^*})}{d\xi} \\ &= (I_{yy} - I_{xx})2s_\xi c_\xi + 2I_{xy}(c_\xi^2 - s_\xi^2) \\ &= (I_{yy} - I_{xx})s_{2\xi} + 2I_{xy}c_{2\xi} \end{aligned}$$

which was obtained by using the relations $\sin 2\xi = 2 \sin \xi \cos \xi$ and $\cos 2\xi = \cos^2 \xi - \sin^2 \xi$. Solving this for ξ we see that the angle ξ for which d^2I attains its maximum satisfies

$$\tan 2\xi = \frac{2I_{xy}}{I_{xx} - I_{yy}}$$

Using Equations (4), (5) and (6), we see that for a sphere

$$\tan 2\xi = \frac{2I_{xy}}{I_{xx} - I_{yy}} = \frac{2\frac{\rho_\lambda}{R}(xy l_3 T^{-3/2})}{\frac{\rho_\lambda}{R}(x^2 l_3 T^{-3/2} - y^2 l_3 T^{-3/2})} = \frac{2xy}{x^2 - y^2}$$

However, noting that for a sphere $\tan r = y/x$ and that $\tan 2r = \frac{2 \tan r}{1 - \tan^2 r}$ we have

$$\tan 2r = \frac{2 \tan r}{1 - \tan^2 r} = \frac{2xy}{x^2 - y^2}$$

Thus, $\xi = r$ and so Proposition 3 is proved.

C. Proposition 5

Let us assume that we have observed that $d^2I = d^2N \cdot L = 0$ in the direction (dx, dy) , and that this situation continues for some distance along (dx, dy) . In this situation, either (1) d^2N is perpendicular to L or (2) the magnitude of d^2N is zero. It is unlikely that d^2N is perpendicular to the illuminant over any distance; thus, if we see that $d^2I = 0$ it must be the case that the magnitude of d^2N is zero. Therefore, dN is some constant vector as we take a step along (dx, dy) .

If the magnitude of dN along (dx, dy) is zero, then at least one of the surface curvatures is zero, i.e., the surface is cylindrical or planar. If the magnitude of dN is not zero, then, when we take a step in the direction (dx, dy) there is some change in surface orientation. In this case, either the amount of foreshortening that occurs with each infinitesimal step along (dx, dy) will change or N will be constant, contrary to the assumption that dN was not zero. If there is change in the foreshortening and yet dN remains constant, the surface orientation relative to the viewer and the intrinsic surface curvature are in an exactly reciprocally relation, which is a violation of general position.

Thus, when $d^2I = 0$ for some distance along a direction (dx, dy) , the surface is either cylindric or planar. If it is planar, $d^2I = 0$ in all directions. The converse — that if we also observe $d^2I = 0$ along the direction orthogonal to (dx, dy) the surface is then planar — is also true. From the previous reasoning, $d^2I = 0$ implies that dN is zero along that direction (if general position is assumed). If dN is zero in two directions, the surface is planar.

D. Proposition 6

Proposition 6 suggests that the following equation is a useful estimator of z_N , the z component of the surface normal, equal to the arccosine of the surface slant:

$$z_N = -c \left(\left| \frac{\nabla^2 I}{I} \right| - c^2 \right)^{-1/2}$$

where c is a constant related to the surface curvature.

We may examine this estimator in the context of the calculations presented so far. First we use Equations (4) and (5) to find that

$$\nabla^2 I = I_{xx} + I_{yy} = \frac{\rho\lambda}{R} (2l_3 T^{-1/2} + (x^2 + y^2)l_3 T^{-3/2})$$

Thus,

$$\frac{\nabla^2 I}{I} = \frac{\frac{\rho\lambda}{R} (2l_3 T^{-1/2} + (x^2 + y^2)l_3 T^{-3/2})}{\frac{\rho\lambda}{R} (-xl_1 - yl_2 - l_3 T^{1/2})} \quad (23)$$

If we assume that $-xl_1 - yl_2$ is zero (as is true on the average, although not necessarily true for any one image point) Equation (23) becomes

$$\begin{aligned} \left| \frac{\nabla^2 I}{I} \right| &= 2T^{-1} + (x^2 + y^2)T^{-2} \\ &= \frac{2R^2 - x^2 - y^2}{(R^2 - x^2 - y^2)^2} \\ &= \frac{R^2 + T}{T^2} \end{aligned} \quad (24)$$

To show how this estimator might work, let us first take $c = T^{-1/2} = (R^2 - x^2 - y^2)^{-1/2}$; we then find that, in fact,

$$\begin{aligned} -c(|\frac{\nabla^2 I}{I}| - c^2)^{-1/2} &= -T^{-1/2} \left(\frac{R^2 + T}{T^2} - \frac{1}{T} \right)^{-1/2} \\ &= -\sqrt{\frac{T}{R^2}} \\ &= -\sqrt{\frac{R^2 - x^2 - y^2}{R^2}} = z_N \end{aligned}$$

The problem with actually using this choice of c is that it requires prior knowledge of the surface slant. We must find a constant for our choice of c . If we take $c = 1/R$, then we find that

$$-c(|\frac{\nabla^2 I}{I}| - c^2)^{-1/2} = -\sqrt{\frac{R^2 - x^2 - y^2}{R^2 - \epsilon}} \quad (25)$$

where

$$\epsilon = R^2 \left(\frac{T}{R^2} - \frac{R^2}{T} \right) = R^2(z_N^2 - z_N^{-2}) \quad (26)$$

Because ϵ is a term in z_N^{-1} the estimated surface slant will be larger than the actual one, introducing a bias into our estimate. This bias can be removed, however, as follows. Let $s = c(|\frac{\nabla^2 I}{I}| - c^2)^{-1/2}$ be the estimated surface slant. Then, combining Equations (25) and (26), we have

$$s^2 = \frac{T}{R^2 - R^2(z_N^2 - z_N^{-2})}$$

Then

$$s^2(1 - z_N^2 + z_N^{-2}) = \frac{T}{R^2} = z_N^2$$

so that we obtain a quadratic in z_N^2

$$(1 + s^2)z_N^4 - s^2 z_N^2 - s^2 = 0 \quad (27)$$

Equation (27) can then be solved to obtain an unbiased estimate of z_N

$$z_N = -\sqrt{\frac{s^2 + \sqrt{4s^2 + 5s^4}}{2 + 2s^2}}$$

(There is only one solution as $0 \geq z_N \geq -1$.) This concludes the proof of Proposition 5.

This estimate of z_N , while unbiased, is not exact because it was necessary to assume that $-xl_1 - yl_2 = 0$. If we examine the conditions under which this factor causes significant error, we see that there will be large errors only when all of the following conditions occur simultaneously:

- (1) The surface slant is relatively large
- (2) The z -component of the illuminant direction, l_3 , is small
- (3) The surface faces closely toward or away from the illuminant.

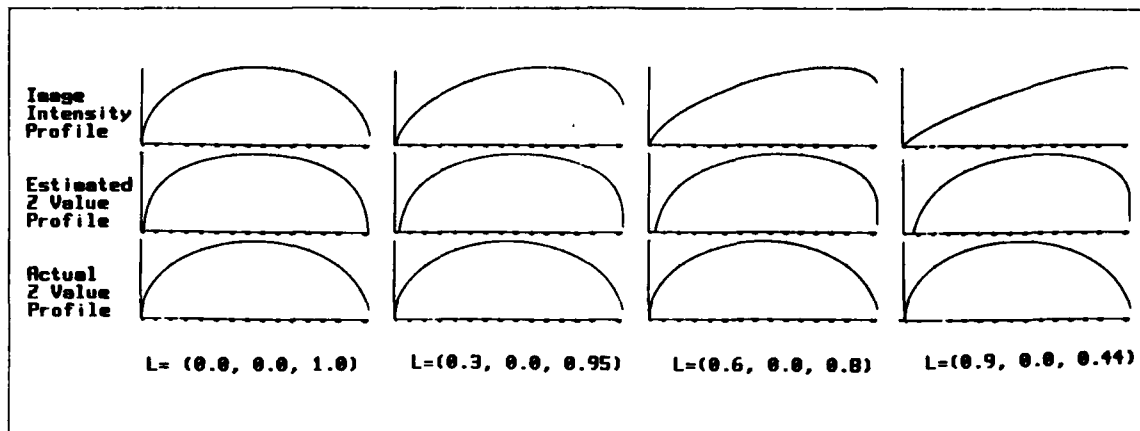


Figure 12. Bias due to illuminant direction. This figure shows the image intensity profile, the profile of the true surface shape (a sphere) and the profile of the reconstructed surface for four illumination conditions. In each case the profile is taken along the image line which goes through the center of the sphere and directly toward the illuminant. This is the direction along which the estimation errors are largest. The distributions of illumination are extended sources, such as would occur if the imaged sphere were placed on a desktop which was near a window. The leftmost distribution shown is centered directly behind the viewer at $(0.0, 0.0, 1.0)$, the next (proceeding left to right) is centered at $(0.3, 0.0, 0.954)$, the next at $(0.6, 0.0, 0.8)$, and the rightmost at $(0.9, 0.0, 0.436)$.

Figure 12 shows the bias due to illuminant direction which occurs during the estimation of surface orientation for the image of a sphere. This figure shows the image intensity profile, the profile of the true surface shape (a sphere) and the profile of the reconstructed surface for four illumination conditions. In each case the profile is taken along the image line which goes through the center of the sphere and directly toward the illuminant. This is the direction along which the estimation errors are largest.

The distributions of illumination are extended sources, such as would occur if the imaged sphere were placed on a desktop which was near a window. The leftmost distribution shown is centered directly behind the viewer at $(0.0, 0.0, 1.0)$, the next (proceeding left to right) is centered at $(0.3, 0.0, 0.954)$, the next at $(0.6, 0.0, 0.8)$, and the rightmost at $(0.9, 0.0, 0.436)$. Note that the rightmost distribution of illumination results in an almost linear gradient across the image. Thus, these examples approximately span the range of illumination directions found in natural scenes.

Comparing the true surface shape, shown across the bottom of Figure 12, to the reconstructed surface shape²¹ shown across the middle of Figure 12, we see that the bias due to illuminant direction does not cause large errors

²¹ As obtained by integrating the estimated surface orientation.

REFERENCES

- [1] A. P. Pentland, "The Visual Inference Of Shape: Computation From Local Features," Ph.D. Thesis, Psychology Department, Massachusetts Institute of Technology (1982)
- [2] B.K.P. Horn, "Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View", A.I. Technical Report 79, Project MAC, Massachusetts Institute of Technology, (1970)
- [3] K. Ikeuchi and B.K.P. Horn, "Numerical Shape from Shading and Occluding Boundaries," Artificial Intelligence, Special Issue on Computer Vision, 15 (1981)
- [4] A. Bruss, "Shape From Shading And Bounding Contour," Ph.D. Thesis, Dept. Elec. Engr. & Comp. Sci., Massachusetts Institute of Technology, (1981).
- [5] A. Pentland "Finding The Illuminant Direction," Journal of the Optical Society of America, Vol 72, No 4 (1982)
- [6] D. Scharf "Magnifications — Photography With The Scanning Electron Microscope," Schocken Brooks, New York, New York
- [7] D. Marr and E. Hildreth, "Theory of edge detection," Massachusetts Institute of Technology AI Memo 518 (1979).
- [8] S. R. Cajal, "Histologie du systeme nerveux de l'homme et des vertebrates," 2, Paris, Maloine (1911).
- [9] S. W. Kuffler, "Discharge Patterns And Functional Organization Of The Mammalian Retina," J. Neurophys. 16 (1953), pp. 281-292.
- [10] D. H. Kelly, "Spatial Frequency Selectivity In The Retina," Vision Research 15 (1975), pp. 665-672.
- [11] D. H. Kelly, "Motion And Vision. II. Stabilized Spatio-Temporal Threshold Surface," Journal of the Optical Society of America 69, 10 (1979), pp. 1340-1349.
- [12] D. H. Hubel and T. N. Wiesel, "The Ferrier Lecture: Functional Architecture Of Macaque Monkey Visual Cortex," R. Soc. Lond. B. 198 (1977), pp. 1-59.
- [13] P. H. Schiller, B. L. Findlay, and S. F. Volman, "Quantitative Studies Of Single Cell Properties In Monkey Striate Cortex I. Spatio-Temporal Organization Of Receptive Fields," J. Neurophys. 39 (1976), pp. 1288-1319.
- [14] K. De Valois, R. L. De Valois, and E. W. Yund, "Responses Of Striate Cortex Cells To Grating And Checkerboard Patterns," J. Physiology 291 (1979) pp. 483-505.
- [15] D. Rose, "Mechanisms Underlying The Receptive Field Properties Of Neurons In Cat Visual Cortex," Vision Research 19 (1979) pp. 533-544.
- [16] J. G. Daugman, "Two-Dimensional Spectral Analysis Of Cortical Receptive Field Profiles," Vision Research 20 (1980) pp. 847-856
- [17] A. P. Witkin, "Recovering Surface Shape and Orientation from Texture," Artificial Intelligence, 17, pp. 17-47 (1981).
- [18] J. R. Kender, "Shape From Texture: An Aggregation Transform that Maps a Class of Textures Into Surface Orientation," Proceedings of the Sixth International Joint Conference on Artificial Intelligence, Tokyo, Japan (1979).
- [19] A. P. Pentland, "Remote Sensing And Human Visual Processing," Proceedings of the Seventh Annual Conference on Remote Sensing of the Environment, San Jose, Costa Rica, (1980).

- [20] C. M. Brown, D. H. Ballard, and O. A. Kimball, "Constraint Interaction In Shape-From-Shading Algorithms," Proceedings of the Image Understanding Workshop, Palo Alto, Ca., (1982).

Appendix B

FROM IMAGE IRRADIANCE TO SURFACE ORIENTATION

Grahame Smith

SRI International



FROM IMAGE IRRADIANCE TO SURFACE ORIENTATION

Technical Note 273

December 1982

By: Grahame B. Smith, Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

The research reported herein was supported by the Defense Advanced Research Projects Agency under Contract Nos. MDA903-79-C-0588 and DAAG29-79-C-0216; these contracts are monitored by the U.S. Army Engineer Topographic Laboratory and by the U.S. Army Research Office.

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 859-6200 • TWX: 910-373-2046 • Telex: 334 486

FROM IMAGE IRRADIANCE TO SURFACE ORIENTATION.

Grahame B. Smith

Artificial Intelligence Center, SRI International
Menlo Park, California 94025

ABSTRACT

The image irradiance equation constrains the relationship between surface orientation in a scene and the irradiance of its image. This equation requires detailed knowledge of both the scene illumination and the reflectance of the surface material. For this equation to be used to recover surface orientation from image irradiance, additional constraints are necessary. The constraints usually employed require that the recovered surface be smooth. We demonstrate that smoothness is not sufficient for this task.

A new formulation of shape from shading is presented in which surface orientation is related to image irradiance without requiring detailed knowledge of the scene illumination, or of the albedo of the surface material. This formulation, which assumes isotropic scattering, provides some interesting performance parallels to those exhibited by the human visual system.

1 INTRODUCTION

Most previous work [1-8] on the problem of recovering surface shape from image shading has been based on solving the image irradiance equation, which relates the radiance of a scene to the irradiance of its image [1,2].¹ This formulation of the relationship between scene radiance and image irradiance is embodied in a first-order partial differential equation expressing scene depth as a function of image coordinates. Such a formulation requires specific knowledge of not only the reflectance characteristics of the surfaces in the scene, but also the position and strength of illumination sources. The approaches to solving this differential equation have generally been either by direct integration [1] or through an iterative algorithm that attempts to reduce the difference between the predicted image irradiance and the measured value [5-7]. Our interest is in the iterative approach because the alternative to it — direct integration — requires specific boundary conditions that are generally unknown (in natural scenes), and its behavior when applied to noisy pictures, is uncertain.

As the image irradiance equation is a single equation relating image irradiance and two independent variables (specifying surface orientation), it does not uniquely determine the two independent variables for a given value of image irradiance. Consequently, when this equation is used to recover surface shape additional constraints are necessary. These may

¹Image irradiance is the light flux per unit area falling on the image, i.e., incident flux density. Scene radiance is the light flux per unit projected area per unit solid angle emitted from the scene, i.e., emitted flux density per unit solid angle.

be imposed by boundary conditions, by restrictions on the type of surface to be recovered, or by a combination of the two. For some images, when we can determine important features (such as the fact that an edge is an occlusion boundary caused by a surface turning smoothly away from the viewing direction), we can use boundary conditions to constrain the solution; in large portions of the image, however, we can say something only about the type of surface we would like to recover. To date surface smoothness is the weakest additional assumption that has allowed surface shape to be recovered. Smoothness normally signifies that the surface is continuous and that it is once or twice differentiable. Smoothness, as the additional assumption, has had to play the role of propagator of boundary conditions and selector of the surface to be recovered. Is smoothness capable of these tasks in general or is its usefulness limited to special cases?

In the first part of this paper we describe the various formulations that have employed smoothness, including a relaxation procedure of our own that resembles its counterpart in engineering; we then present results of our experiments with these iterative procedures. Assessing the usefulness of smoothness in this context, we conjecture as to its utility in other shape-from-shading formulations.

Not all authors have used smoothness as their additional constraint; some have employed assumptions about surface shape instead. The assumption that the surface is locally spherical, i.e., that its curvature is independent of direction, is strong enough to allow but a single interpretation for the surface orientation, and at the same time, it is also one that enables recovery of the surface orientation by purely local computation [8]. In addition, this shape constraint eliminates the need to know such parameters as illuminant direction and surface albedo.² Assumptions about shape are being traded for assumptions about reflectance behavior. Can we formulate the shape-from-shading problem without having to know the details of the surface reflectance and without making any assumptions about the shape of the surface we wish to recover?

In the new formulation presented in the second part of the paper, we assume that scene materials scatter light isotropically. We make no assumptions about surface shape and we do not need to know the parameters specifying illuminant direction, illuminant strength, and surface albedo. Our assumptions are about the properties of reflection in the world; these alone are sufficient to relate surface orientation to image irradiance. In situations in which the assumption of isotropic scattering is invalid, the formulation provides some interesting parallels to human vision.

2 ITERATIVE FORMULATIONS FOR SURFACE RECOVERY

The image irradiance equation as presented by Horn [2], is

$$I(x, y) = R(p, q) \quad ,$$

where $I(x, y)$ is the image irradiance as a function of the image coordinates x and y , and $R(p, q)$ is the surface radiance as a function of p and q , the derivatives of depth with respect

²Surface albedo is the material reflectance, i.e., the ratio of scene radiance to scene irradiance.

to the image coordinates. To derive this equation orthographic projection is assumed; while orthographic projection is inadequate to describe image formation it is a good approximation when the scene objects are small compared to the viewing distance. In the shape-from-shading approach, it is generally assumed that $R(p, q)$ is known for all p and q (that is, the reflectance map is specified). The iterative approach applies this equation on a pixel-by-pixel basis, that is, for pixel (i, j)

$$I_{i,j} = R(p_{i,j}, q_{i,j}) \quad ,$$

where $I_{i,j}$ is the image irradiance for (i, j) th pixel and $p_{i,j}, q_{i,j}$ is the surface orientation of the surface patch that is imaged at pixel (i, j) . For convenience we use the notation

$$R_{i,j} \equiv R(p_{i,j}, q_{i,j}) \quad .$$

If, at some stage of the iterative procedure, we have assigned particular $p_{i,j}, q_{i,j}$ as the surface orientation of the (i, j) th pixel, then the residual expression

$$\xi_{i,j}^R = (I_{i,j} - R_{i,j})^2$$

specifies the error caused by our assignment of surface orientation.³ If this were our only constraint, we could select $p_{i,j}, q_{i,j}$ so that $\xi_{i,j}^R = 0$. This would guarantee that the image irradiance equation is satisfied pixel by pixel, but, because there are infinitely many solutions, we need further constraints to reduce the number of possible solutions.

Smoothness is usually introduced by specifying a relationship that we would like to have hold between the surface orientation of the (i, j) th pixel and its neighbors. The various iterative approaches [5-7] differ in the way this relationship is specified. Of course, at a particular stage of the iterative process this relationship between a pixel and its neighbors will not be exact. Once again we can specify a residual equation for the error in the smoothness relation.

$$\xi_{i,j}^S = [f(p_{i,j}, q_{i,j}, p_{i-1,j}, q_{i-1,j}, p_{i+1,j}, q_{i+1,j}, p_{i,j-1}, q_{i,j-1}, p_{i,j+1}, q_{i,j+1}, \dots)]^2 \quad ,$$

where f is the relationship between the surface orientation at (i, j) and its neighbors. An example of the type of relationship is the difference between the surface orientation of pixel (i, j) and the mean value of the surface orientations of its four-neighbors.

We have two constraints that need to be satisfied simultaneously, — one from image irradiance and one from surface smoothness. At each stage of the iterative process, the total residual error for pixel (i, j) can be described by

$$\xi_{i,j} = \lambda \xi_{i,j}^R + \xi_{i,j}^S \quad ,$$

where λ is a weighting factor that can adjust the influence of the error in image irradiance to the error in smoothness.⁴ For the image, the total residual error is

$$\xi = \sum_{i,j} \xi_{i,j} \quad .$$

³The form of the error need not be quadratic — the goals of such a choice include simple final expressions.

⁴Since the error in image irradiance is not necessarily commensurate with that in surface smoothness, some form of normalization is required.

The allocation of surface orientations to all pixels should minimize this total error, that is,

$$\frac{\partial \xi}{\partial p_{i,j}} = 0 \quad ,$$

$$\frac{\partial \xi}{\partial q_{i,j}} = 0 \quad .$$

Differentiating ξ with respect to $p_{i,j}$ and also with respect to $q_{i,j}$ gives two equations for each pixel in the image. While complicated forms of the relationship between $p_{i,j}$, $q_{i,j}$ and their neighboring pixels will generally occur, we choose our smoothness relation so that we can arrange the equations in open form

$$p_{i,j} = F_1(p_{i,j}, q_{i,j}, \text{and } p\text{'s and } q\text{'s of neighboring pixels}) \quad ,$$

$$q_{i,j} = F_2(p_{i,j}, q_{i,j}, \text{and } p\text{'s and } q\text{'s of neighboring pixels}) \quad ,$$

where F_1 , and F_2 are functions.

We therefore have an iterative scheme that, given some initial solution, we improve by reducing the residual error in image irradiance and surface smoothness. We need to ask the following questions of such a scheme: Under what conditions will it converge to a solution? Is that solution unique? Does smoothness, as defined by our relation, give us the type of surface we want?

3 SURFACE ORIENTATION

There are many equivalent parameterizations of surface orientation. Mentioned previously were the parameters p and q , the derivatives of depth with respect to image coordinates. Some authors prefer using slant and tilt to specify surface orientation. Slant is the angle between the surface normal and the viewing direction, while tilt is the angle between the image x axis and the projection of the surface normal onto the image plane. Other parameterizations [7] have been used when particular properties of the parameterization are to be exploited. The parameters we use are l and m :⁵

$$\begin{aligned} l &= \sin \sigma \cos \tau \quad , \\ m &= \sin \sigma \sin \tau \quad , \end{aligned}$$

where σ is the surface slant and τ its tilt, l is the component of the surface normal in the direction of the x axis, and m is the component in the y direction. We select this particular parameterization, as l and m are bounded

$$0 \leq l^2 + m^2 \leq 1 \quad .$$

For surfaces that we can see,

$$0 \leq \sigma < \frac{\pi}{2} \quad ,$$

⁵ l and m are related to p and q . $l = \frac{-p}{\sqrt{1+p^2+q^2}}$, $m = \frac{-q}{\sqrt{1+p^2+q^2}}$.

$$0 \leq \tau < 2\pi$$

Consequently l and m specify the surface normal of an imaged surface without ambiguity.

4 FORMULATIONS USED FOR SURFACE RECOVERY

To explore the issues of convergence, propagation of boundary conditions, and the type of surface promoted by smoothness, we formulate the problem in two ways: one that parallels the technique previously described and, alternatively, one that resembles the relaxation method used to solve structural engineering problems.

The function for scene radiance, used to create synthetic images for the experiment and employed by the shape recovery algorithms, is

$$R(l, m) = 0.1569 \left(\frac{1 + \sqrt{1 - l^2 - m^2}}{2} \right) + \text{Max}[0.4437\sqrt{1 - l^2 - m^2} + 0.3137l + 0.3137m, 0].$$

This function is appropriate for a scene that exhibits Lambertian reflectance and is illuminated by both a collimated source and a uniform hemispherical source. This illumination was selected because it is typical of the illumination of outdoor scenes. The particular numerical constants specify the light direction and intensity, and the surface albedo.

The first formulation is similar to that described previously; we shall call this the 'conventional' formulation. From the image irradiance equation we have the error term

$$\xi_{i,j}^R = (I_{i,j} - R_{i,j})^2$$

The smoothness constraint is the requirement that $l_{i,j}$ be the average of its four-neighbors, and that $m_{i,j}$ be the average of its four-neighbors. The error term for smoothness is

$$\xi_{i,j}^S = \left(l_{i,j} - \frac{l_{i-1,j} + l_{i+1,j} + l_{i,j-1} + l_{i,j+1}}{4} \right)^2 + \left(m_{i,j} - \frac{m_{i-1,j} + m_{i+1,j} + m_{i,j-1} + m_{i,j+1}}{4} \right)^2$$

Note that this constraint is exact for a surface that is spherical.

Minimizing $\xi = \sum_{i,j} \lambda \xi_{i,j}^R + \xi_{i,j}^S$ by differentiating with respect to $l_{i,j}$, and with respect to $m_{i,j}$, and then setting each result equal to zero, we obtain the expressions

$$\begin{aligned} l_{i,j} = & 0.4(l_{i-1,j} + l_{i+1,j} + l_{i,j-1} + l_{i,j+1}) - \\ & 0.1(l_{i-1,j-1} + l_{i+1,j+1} + l_{i-1,j+1} + l_{i+1,j-1}) - \\ & 0.05(l_{i-2,j} + l_{i+2,j} + l_{i,j-2} + l_{i,j+2}) + \\ & 0.8\lambda(I_{i,j} - R_{i,j}) \frac{\partial R}{\partial l} \Big|_{i,j}, \\ m_{i,j} = & 0.4(m_{i-1,j} + m_{i+1,j} + m_{i,j-1} + m_{i,j+1}) - \\ & 0.1(m_{i-1,j-1} + m_{i+1,j+1} + m_{i-1,j+1} + m_{i+1,j-1}) - \\ & 0.05(m_{i-2,j} + m_{i+2,j} + m_{i,j-2} + m_{i,j+2}) + \\ & 0.8\lambda(I_{i,j} - R_{i,j}) \frac{\partial R}{\partial m} \Big|_{i,j}. \end{aligned}$$

We use these (together with the expression for $R(l, m)$) as our iterative scheme to improve on an initial solution.

The other formulation we use, the 'engineering' formulation, creates error terms from the image irradiance equation and the smoothness constraints, but does not combine these into one term.

$$\begin{aligned}\xi_{i,j}^R &= (I_{i,j} - R_{i,j}) \quad , \\ \xi_{i,j}^{S_1} &= (l_{i,j} - \frac{l_{i-1,j} + l_{i+1,j} + l_{i,j-1} + l_{i,j+1}}{4}) \quad , \\ \xi_{i,j}^{S_2} &= (m_{i,j} - \frac{m_{i-1,j} + m_{i+1,j} + m_{i,j-1} + m_{i,j+1}}{4}) \quad .\end{aligned}$$

We view the ξ 's as residuals and apply the relaxation approach, i.e., reduction of the largest residuals. If $\xi_{i,j}^{S_1}$ or $\xi_{i,j}^{S_2}$ is selected for reduction we choose to reduce both, as each is independent of the other. When $\xi_{i,j}^R$ is chosen for reduction, we do the reduction in two stages — one stage altering $l_{i,j}$ and the other $m_{i,j}$. Of course we can scale the residuals, reduce them from, say, the image irradiance equation to a certain level before introducing smoothness, vary the amount of correction we apply, (e.g., we can overrelax) and the like. In fact, we can experiment with various relaxation approaches. In this formulation major changes in the relaxation scheme generally require minor programming changes.

5 EXPERIMENTAL RESULTS

The test image shown in Figure 1 is that of a hemisphere placed on a plane, i.e., a synthetic image generated by the reflectance function previously described. The collimated light source is at slant $\frac{\pi}{4}$ and tilt $\frac{\pi}{4}$ — which means that it is at the upper right as we view the image. We purposely avoided the case in which the collimated source is at the same position as the viewer, since the resulting symmetric reflectance map might bias the algorithm to return a symmetric surface. A synthetic image of a sphere was selected as the test image because both the image irradiance equation and the smoothness relationship we use hold exactly.⁶ The performance of the algorithm to recover the surface shape could be assessed without the complications involved in using inexact models for reflectance and smoothness.

We need initial solutions to start our iterative/relaxation procedures. We used four sets of initial conditions: (1) a plane perpendicular to the viewing direction; (2) a plane slanted $\frac{\pi}{4}$ to the viewing direction; (3) a cone with its axis along the viewing direction; (4) the correct solution perturbed by small random errors.

Previous work has used boundary conditions to constrain the recovered surface. Investigating this approach, we constrained the surface in various ways: at the edge of the hemisphere, at a closed curve lying on the sphere's surface, or at individual points on the sphere's surface. We also used the algorithms without any boundary conditions whatsoever.

Since we wished to investigate the extent to which smoothness could propagate boundary conditions, we used various image quantizations, namely 16×16 , 32×32 , and 64×64 .

⁶The smoothness relationship does not hold at the edge of the hemisphere where it joins the plane.

The findings can be characterized as follows:

- Both techniques — the engineering and the conventional method — gave essentially the same results.
- The engineering technique converged much faster than the conventional technique.
- Smoothness propagates boundary conditions by no more than a few pixels
- The initial solution largely predetermines the final one.

Figures 3-11 display examples of the results we achieved with the conventional iterative scheme; the engineering scheme gave essentially the same results. In each of these figures the top left picture shows the profile of the recovered surface (viewed from the bottom left corner), while at the top right we find an image that is the sine of the surface slant, with black representing 0 and white 1. The bottom left is the cosine and the bottom right the sine of the surface tilt, with black representing -1, gray 0, and white +1. The results are presented in this manner so that the performance of the algorithms can be evaluated. The profile can on occasion appear more accurate than the individual surface orientations (as might be expected of an integration procedure); at other times, however, errors in the surface orientation (sometimes just from the image quantization) of highly slanted surfaces cause the integration routine that produces the surface shape for profiling to overstate the error. Figure 2 shows the results that should be obtained if the shape recovery algorithms recovered the surface exactly.

Figures 3-6 illustrate the effects of various boundary conditions. The errors at the edge of the sphere where it joins the plane are expected, as smoothness does not hold there. Each figure is the result of 320 iterations, this being five times the linear dimension of the picture used. The boundary condition at a point affects an area of approximately 10 pixels in radius. Only for Figure 6, where a random five percent of pixels were set to their correct values, is the surface shape recovered correctly. Smoothness as a propagator affects but a small area. Figures 4, 7, and 8 illustrate this point further. Here various image sizes are used. Observe that, as the image size increases, the boundary conditions diminish in their effect and the solution becomes progressively worse. Figures 4, 9, and 10 reveal the dominant influence of the initial solution. Figure 11 is included to show the effect of smoothness when $\lambda = 0$ — namely, when image irradiance does not affect the solution at all. This figure, obtained after 320 iterations, demonstrates what smoothness alone can achieve, even when the definition of smoothness is exact for the viewed scene (a sphere).

Smoothness is a poor selector of surface shape and a poor propagator of boundary information when it is used to tie the surface orientation of a particular surface point to those of its neighbors. Generally, in engineering, problems solved with relaxation techniques are formulations that relate a given property at one point to that same property at neighboring points by means of differential relations. It is the derivative that propagates boundary information and selects a particular solution to be recovered. We present such a formulation below in an attempt to relieve smoothness of its role as propagator and selector.

6 SURFACE RADIANCE AND ISOTROPIC SCATTERING

Our formulation of the relationship between image irradiance and scene radiance is

$$I(x, y) = R(l, m) ,$$

where $I(x, y)$ is the image irradiance at image point x, y and $R(l, m)$ is the scene radiance for a surface normal we represent by l, m . R is a function of the components of the surface normal and they, in turn, are functions of image coordinates. $R(l, m)$ specifies the relationship between surface radiance and surface orientation, while $l(x, y)$ and $m(x, y)$ specify the relationship between surface orientation and image coordinates. $R(l, m)$ embodies knowledge of the nature of surface reflection, while $l(x, y)$ and $m(x, y)$ embody the surface shape.

To provide the additional constraints we need for relating surface orientation to image irradiance, we introduce constraints that relate properties of $R(l, m)$, — that is, constraints that specify the relationship between surface radiance and surface orientation. Such constraints are

$$\begin{aligned} (1 - l^2)R_{ll} &= (1 - m^2)R_{mm} , \\ (R_{ll} - R_{mm})lm &= (l^2 - m^2)R_{lm} , \end{aligned}$$

where R_{ll} is the second partial derivative of R with respect to l , R_{mm} is the second partial derivative of R with respect to m , and R_{lm} is the second partial cross-derivative of R with respect to l and m .

These two partial differential equations embody the assumption of isotropic scattering (Lambertian reflectance). For isotropic scattering $R(l, m)$ has the form

$$R(l, m) = al + bm + c\sqrt{1 - l^2 - m^2} + d ,$$

where a, b, c , and d are constants, their values depending on illumination conditions and surface albedo. Note that l, m , and $\sqrt{1 - l^2 - m^2}$ are the components of the unit surface normal in the directions x, y , and depth. $R(l, m)$ can be viewed as the dot product of the surface normal vector $(l, m, \sqrt{1 - l^2 - m^2})$ and a vector (a, b, c) denoting illumination conditions. As the value of a dot product is rotationally independent of the coordinate system, the scene radiance is independent of the viewing direction — which is the definition of isotropic scattering.

It is easily seen that $R(l, m) = al + bm + c\sqrt{1 - l^2 - m^2} + d$ satisfies the pair of partial differential equations given above. In the appendix we show that $R(l, m) = al + bm + c\sqrt{1 - l^2 - m^2} + d$ is the solution of the pair of partial differential equations. These partial differential equations are an alternative definition of isotropic scattering.

It is worthy of note that $R(l, m) = al + bm + c\sqrt{1 - l^2 - m^2} + d$ includes radiance functions for multiple and extended illumination sources, including that for a hemispherical uniform source such as the sky. The assumption of isotropic scattering restricts the class of material surfaces being considered, not the illumination conditions.

7 EQUATIONS RELATING SURFACE ORIENTATION TO IMAGE IRRADIANCE

Differentiating

$$I(x, y) = R(l, m)$$

with respect to x and y , we obtain

$$I_x = R_l l_x + R_m m_x ,$$

$$I_y = R_l l_y + R_m m_y ,$$

$$I_{xx} = R_{ll} l_x^2 + R_{mm} m_x^2 + 2R_{lm} l_x m_x + R_l l_{xx} + R_m m_{xx} ,$$

$$I_{yy} = R_{ll} l_y^2 + R_{mm} m_y^2 + 2R_{lm} l_y m_y + R_l l_{yy} + R_m m_{yy} ,$$

$$I_{xy} = R_{ll} l_x l_y + R_{mm} m_x m_y + R_{lm} (l_x m_y + l_y m_x) + R_l l_{xy} + R_m m_{xy} ,$$

where subscripted variables denote partial differentiation with respect to the subscript(s).

From the constraints for isotropic scattering, we derive the relationships

$$R_{ll} = \frac{1 - m^2}{lm} R_{lm} ,$$

$$R_{mm} = \frac{1 - l^2}{lm} R_{lm} .$$

Substituting these relationships for R_{ll} and R_{mm} in the expressions for I_{xx} , I_{yy} , and I_{xy} , we obtain

$$[l_x^2 (\frac{1 - m^2}{lm}) + m_x^2 (\frac{1 - l^2}{lm}) + 2l_x m_x] R_{lm} = I_{xx} - R_l l_{xx} - R_m m_{xx} ,$$

$$[l_y^2 (\frac{1 - m^2}{lm}) + m_y^2 (\frac{1 - l^2}{lm}) + 2l_y m_y] R_{lm} = I_{yy} - R_l l_{yy} - R_m m_{yy} ,$$

$$[l_x l_y (\frac{1 - m^2}{lm}) + m_x m_y (\frac{1 - l^2}{lm}) + l_x m_y + l_y m_x] R_{lm} = I_{xy} - R_l l_{xy} - R_m m_{xy} .$$

By removing R_{lm} and substituting the expressions for R_l and R_m , defined by the expressions for I_x and I_y , we produce two partial differential equations relating surface orientation to image irradiance:

$$\alpha \theta l_{xx} + \beta \theta m_{xx} - \alpha \gamma l_{xy} - \beta \gamma m_{xy} = \chi \theta I_{xx} - \chi \gamma I_{xy} ,$$

$$\alpha \theta l_{yy} + \beta \theta m_{yy} - \alpha \delta l_{xy} - \beta \delta m_{xy} = \chi \theta I_{yy} - \chi \delta I_{xy} ,$$

where

$$\alpha = I_x m_y - I_y m_x ,$$

$$\beta = I_y l_x - I_x l_y ,$$

$$\gamma = l_x^2 (1 - m^2) + m_x^2 (1 - l^2) + 2l_x m_x lm ,$$

$$\delta = l_y^2 (1 - m^2) + m_y^2 (1 - l^2) + 2l_y m_y lm ,$$

$$\theta = l_x l_y (1 - m^2) + m_x m_y (1 - l^2) + (l_x m_y + l_y m_x) lm ,$$

$$\chi = l_x m_y - l_y m_x .$$

These equations relate surface orientation to image irradiance by parameter-free expressions. They involve the derivatives of image irradiance, but not the image irradiance itself — an important feature if we conjecture such a model for the human visual system.

8 RECOVERY OF SURFACE SHAPE — A SPECIAL CASE: A SPHERICAL SURFACE

It is difficult to solve the equations relating surface orientation to image irradiance, and thus to recover surface shape from observed image irradiance. Two types of approaches are possible. The two differential equations can be integrated in a step-by-step manner or, given some initial solution, a relaxation procedure may be employed. The difficulties that arise are two-fold, numerical errors and multiple solutions.

Solutions of the equation $\chi = 0$ (the developable surfaces, e.g., a cylinder) are also solutions of the equations relating surface orientation to image irradiance. If the image intensities were known in analytic form then analytic solution of the equations could employ boundary conditions to select the appropriate solution. However since the analytic form for the image intensities is unknown, numerical procedures must be employed. Numerical procedures to integrate the equations inevitably introduces small errors. Instability of the numerical scheme seems responsible for such errors eventually dominating the recovered solution.

The alternative, a relaxation procedure to solve the equations, has its own difficulties. The difficulties experienced in the shape-from-shading methods discussed in the first part of this paper dictate caution. The importance of a good initial solution for a relaxation method cannot be overemphasized. Simplifying the two partial differential equations (using additional assumptions) provides a method for obtaining an good initial solution.

The spherical approximation assumes that we are on a spherical surface. This implies $l_y = 0, m_x = 0, l_x = m_y$, --- namely, constant curvature independent of direction. For this case the partial differential equations become relationships between the second derivatives of image irradiance and the components of the surface normal:

$$\frac{1 - m^2}{lm} = \frac{I_{xx}}{I_{xy}},$$
$$\frac{1 - l^2}{lm} = \frac{I_{yy}}{I_{xy}}.$$

These results for the spherical approximation are equivalent to those Pentland was able to obtain [8] through local analysis of the surface. In addition to providing a mechanism for obtaining an initial solution for a relaxation-style algorithm, their direct application estimates the surface orientation by local computation [8].

We are actively engaged in the development of a relaxation procedure to transform the initial solution (given by the spherical approximation) into a solution the satisfies the full equations.

9 THE INFLUENCE OF BELIEF ON THE PERFORMANCE OF A VISUAL SYSTEM

The constraints derived for isotropic scattering do not have to be true embodiments of the physical laws of nature; rather, they can represent the beliefs a visual system possesses regarding those laws. In circumstances in which such beliefs do not hold, the visual system

will err in predicting the world's true nature. Of course, if the model is not a good approximation of the physical laws of nature, the visual system embodying it is useless.

The two constraints specifying isotropic scattering,

$$\begin{aligned}(1 - l^2)R_{ll} &= (1 - m^2)R_{mm} \quad , \\ (R_{ll} - R_{mm})lm &= (l^2 - m^2)R_{lm} \quad ,\end{aligned}$$

obviously both hold when the scattering is isotropic, but what is the situation for other forms of scattering?

The images produced by a scanning electron microscope constitute an intriguing situation. The appropriate expression for scene radiance [7] is

$$R(l, m) = a \left(1 + \frac{1}{\sqrt{1 - l^2 - m^2}} \right) \quad ,$$

where a is a constant. This expression is quite unlike those for natural scenery, yet the human visual system 'sees' an image. Note that the second constraint for the isotropic scattering case is satisfied by this radiance function, but not the first. The second constraint is about surface tilt, as $\frac{lm}{l^2 - m^2} = \frac{\tan^2 \tau}{2}$, where τ is the surface tilt; the first constraint introduces slant. In using the equations relating surface orientation to image irradiance to recover surface orientation, one might expect them to predict tilt correctly for surfaces in electron microscope images, but to err in predicting slant.

For other forms of the scene radiance expressions, neither constraint holds. Specular reflectance has been approximated [2] by

$$R(l, m) = a[b(1 - l^2 - m^2) + cl\sqrt{1 - l^2 - m^2} + dm\sqrt{1 - l^2 - m^2}]^n \quad ,$$

where n is a constant, usually having a value between 1 and 10 that determines the 'sharpness' of the specular peak.

For the maria of the moon, the form of scene radiance normally used [2] is

$$R(l, m) = \frac{a(bl + cm)}{\sqrt{1 - l^2 - m^2}} \quad .$$

The constants a, b, c , and d in the above expressions are associated with the strength and position of the light source, as well as with the surface albedo.

The constraints do not hold in either of the preceding cases. We would expect a visual system embodying them to make errors under these circumstances. Nevertheless this should not induce us to immediately begin searching for new visual beliefs. After all the human visual system is imperfect under conditions of specular reflection; moreover, people observed the moon throughout history without concluding that it was spherical.

If these constraints are incorporated in the human visual system, the predictions based on them — i.e., when the visual system will return ostensibly 'correct' and 'incorrect' information — could be tested by psychophysical experiments. Such predictions together with their verification or refutation are being investigated.

10 CONCLUSION

The shape-from-shading task (recovering surface orientation from image irradiance), has meant finding a solution to the image irradiance equation. This formulation requires that the characteristics of the scene illumination and the surface material be known. While these requirements are difficult to satisfy, knowing them makes it possible to apply the image irradiance equation to any scene material for which the scene radiance function is known. Such application, however, is not without difficulty, appropriate boundary conditions are needed and the effect of image noise is uncertain.

To recover surface orientation, relaxation-style algorithms based on the image irradiance equation employ additional constraints. These constraints, which are needed to supplement the underdetermined image irradiance equation, capture the concept of smoothness. While smoothness superficially determines the relationship between image irradiance and surface orientation, it is too weak a concept to propagate boundary conditions and thus equally ineffectual as a means of recovering the required solution.

In presenting a new formulation for the shape-from-shading task, we have traded the need to know the explicit form of the scene radiance function for the assumption that material scatters light isotropically. This model is applicable to natural scenery without additional assumptions about illumination conditions or the albedo of the surface material. The model also demonstrates some competence even when the scattering is not isotropic. Such a model poses the question: does the human visual system embody a particular belief about the laws of scattering that it applies even when these laws are inexact?

Effective numerical procedures based on this new formulation of the shape-from-shading task remain unknown and, are subjects for further development.

ACKNOWLEDGMENT

The research reported herein was supported by the Defense Advanced Research Projects Agency under Contracts MDA903-79-C-0588 and DAAG29-79-C-0216; these contracts are monitored by the U.S. Army Engineer Topographic Laboratory and by the U.S. Army Research Office.

REFERENCES

1. Horn, B.K.P., Obtaining shape from shading information, in: Winston, P.H. (Ed.), *Psychology of Computer Vision*, McGraw-Hill, New York, 1975 115-155.
2. Horn, B.K.P., Understanding image intensities, *Artificial Intelligence* 8 (1977) 201-231.
3. Woodham, R.J., A cooperative algorithm for determining surface orientation from a single view, *Proceedings of the Fifth IJCAI*, Cambridge, Massachusetts, 1977 635-641.
4. Woodham, R.J., Relating properties of surface curvature to image intensity, *Proceedings of the Sixth IJCAI*, Tokyo, Japan, 1979 971-977.
5. Brooks, M.J., Surface normals from closed paths, *Proceedings of the Sixth IJCAI*,

Tokyo, Japan, 1979 98-101.

6. Strat, T.M., A numerical method for shape from shading from a single image, S.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (1979).
7. Ikeuchi, K. and Horn, B.K.P., Numerical shape from shading and occluding boundaries, *Artificial Intelligence* 17 (1981) 141-184.
8. Pentland, A.P., The visual inference of shape: computation from local features, Ph.D. Thesis, Department of Psychology, Massachusetts Institute of Technology (1982).

APPENDIX

We show that the solution of the system of partial differential equations,

$$\begin{aligned}(1 - l^2)R_{ll} &= (1 - m^2)R_{mm} \quad , \\ (R_{ll} - R_{mm})lm &= (l^2 - m^2)R_{lm} \quad ,\end{aligned}$$

where R_{ll} is the second partial derivative of R with respect to l , R_{mm} is the second partial derivative of R with respect to m , and R_{lm} is the second partial cross derivative of R with respect to l and m , is

$$R(l, m) = al + bm + c\sqrt{1 - l^2 - m^2} + d \quad ,$$

where a, b, c , and d are constants.

Proof: Rearranging

$$\begin{aligned}(1 - l^2)R_{ll} &= (1 - m^2)R_{mm} \quad , \\ (R_{ll} - R_{mm})lm &= (l^2 - m^2)R_{lm} \quad ,\end{aligned}$$

we obtain

$$\begin{aligned}R_{lm} &= \frac{lm}{1 - m^2} R_{ll} \quad , \\ R_{lm} &= \frac{lm}{1 - l^2} R_{mm} \quad .\end{aligned}$$

Integrating $R_{lm} = \frac{lm}{1 - m^2} R_{ll}$ with respect to l , that is,

$$\int R_{lm} dl = \frac{m}{1 - m^2} \int l R_{ll} dl \quad ,$$

gives

$$R_m = \frac{m}{1 - m^2} (l R_l - R) + F_1(m) \quad ,$$

where $F_1(m)$ is an arbitrary function of m . Similarly, integrating $R_{lm} = \frac{lm}{1 - l^2} R_{mm}$ with respect to m gives

$$R_l = \frac{l}{1 - l^2} (m R_m - R) + F_2(l) \quad ,$$

where $F_2(l)$ is an arbitrary function of l . Rearranging these two equations, we get a system of two first order partial differential equations

$$(1 - l^2 - m^2)R_l + lR = lmF_3(m) + (1 - m^2)F_4(l) ,$$

$$(1 - l^2 - m^2)R_m + mR = lmF_4(l) + (1 - l^2)F_3(m) ,$$

where $F_3(m) = (1 - m^2)F_1(m)$, and $F_4(l) = (1 - l^2)F_2(l)$. Multiplying both equations by the integrating factor $(1 - l^2 - m^2)^{-\frac{3}{2}}$, we obtain

$$\frac{\partial}{\partial l}[(1 - l^2 - m^2)^{-\frac{1}{2}}R] = (1 - l^2 - m^2)^{-\frac{3}{2}}[lmF_3(m) + (1 - m^2)F_4(l)] ,$$

$$\frac{\partial}{\partial m}[(1 - l^2 - m^2)^{-\frac{1}{2}}R] = (1 - l^2 - m^2)^{-\frac{3}{2}}[lmF_4(l) + (1 - l^2)F_3(m)] .$$

Before carrying out the integration, we can find the form of $F_3(m)$, and $F_4(l)$ by differentiating the first equation with respect to m and the second with respect to l :

$$\frac{\partial^2}{\partial l \partial m}[(1 - l^2 - m^2)^{-\frac{1}{2}}R] = (1 - l^2 - m^2)^{-\frac{5}{2}}[l(1 - l^2 + 2m^2)F_3(m) + m(1 - m^2 + 2l^2)F_4(l) + lm(1 - l^2 - m^2)F'_3(m)] ,$$

$$\frac{\partial^2}{\partial l \partial m}[(1 - l^2 - m^2)^{-\frac{1}{2}}R] = (1 - l^2 - m^2)^{-\frac{5}{2}}[l(1 - l^2 + 2m^2)F_3(m) + m(1 - m^2 + 2l^2)F_4(l) + lm(1 - l^2 - m^2)F'_4(l)] ,$$

where $F'(k)$ indicates differentiation with respect to the independent variable k . Hence,

$$F'_3(m) = F'_4(l) .$$

$F_3(m)$ is a function of m and $F_4(l)$ is a function of l ; this implies that

$$F'_3(m) = d ,$$

$$F'_4(l) = d ,$$

where d is a constant. Therefore,

$$F_3(m) = dm + b ,$$

$$F_4(l) = dl + a ,$$

where a , and b are constants. Returning to the integration step, we now have the expressions

$$\frac{\partial}{\partial l}[(1 - l^2 - m^2)^{-\frac{1}{2}}R] = (1 - l^2 - m^2)^{-\frac{3}{2}}[l(bm + d) + a(1 - m^2)] ,$$

$$\frac{\partial}{\partial m}[(1 - l^2 - m^2)^{-\frac{1}{2}}R] = (1 - l^2 - m^2)^{-\frac{3}{2}}[m(al + d) + b(1 - l^2)] .$$

Integrating the first equation with respect to l and the second with respect to m , we obtain

$$(1 - l^2 - m^2)^{-\frac{1}{2}}R = (bm + d)(1 - l^2 - m^2)^{-\frac{1}{2}} + al(1 - l^2 - m^2)^{-\frac{1}{2}} + F_5(m) ,$$

$$(1 - l^2 - m^2)^{-\frac{1}{2}}R = (al + d)(1 - l^2 - m^2)^{-\frac{1}{2}} + bm(1 - l^2 - m^2)^{-\frac{1}{2}} + F_6(l) ,$$

where $F_5(m)$, and $F_6(l)$ are arbitrary functions of m and l , respectively. We have two expressions for R :

$$R = al + bm + F_5(m)(1 - l^2 - m^2)^{\frac{1}{2}} + d ,$$

$$R = al + bm + F_6(l)(1 - l^2 - m^2)^{\frac{1}{2}} + d ,$$

which are compatible if

$$F_5(m) = F_6(l) = c ,$$

where c is a constant. The solution for R is

$$R = al + bm + c\sqrt{1 - l^2 - m^2} + d .$$

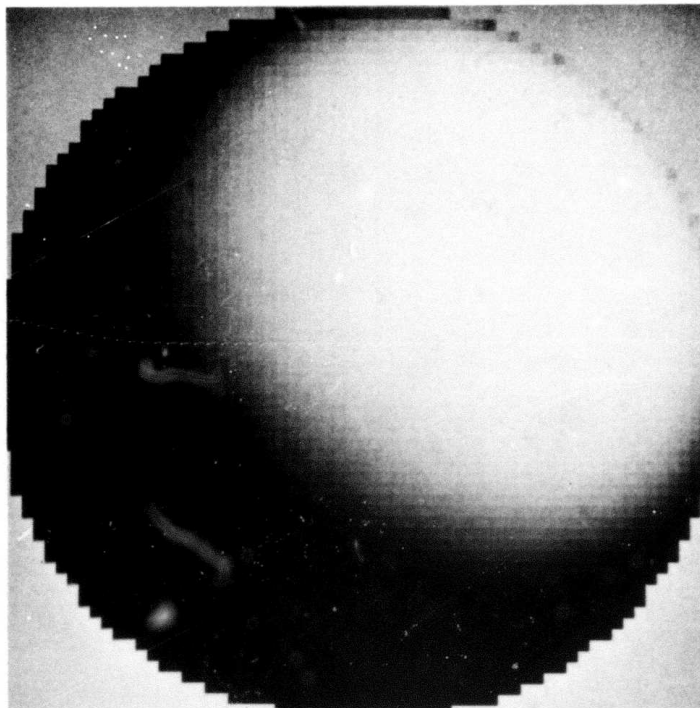


Figure 1 Original Image.

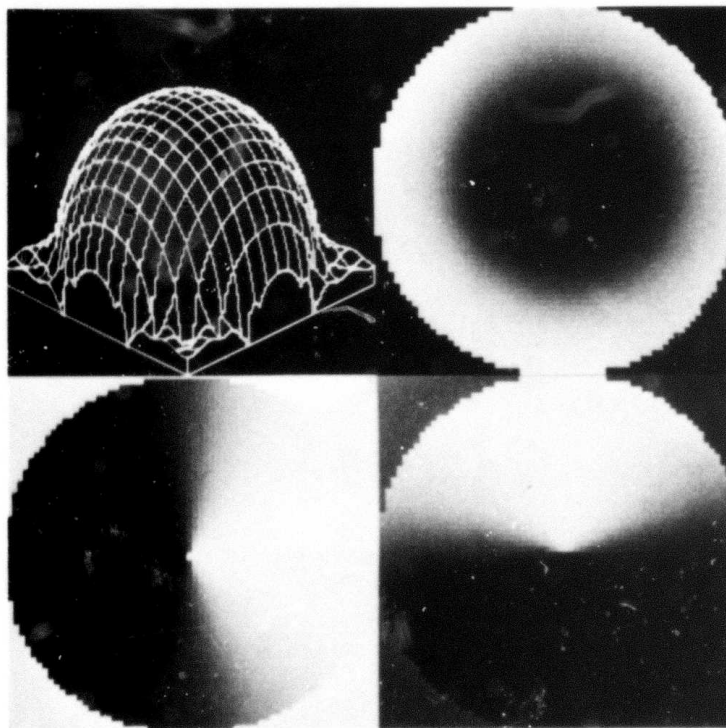


Figure 2 'Ideal' Result. Top left - profile of recovered surface; top right - sine slant, black \equiv 0, white \equiv 1; bottom left - cosine tilt, black \equiv -1, gray \equiv 0, white \equiv +1; bottom right - sine tilt.

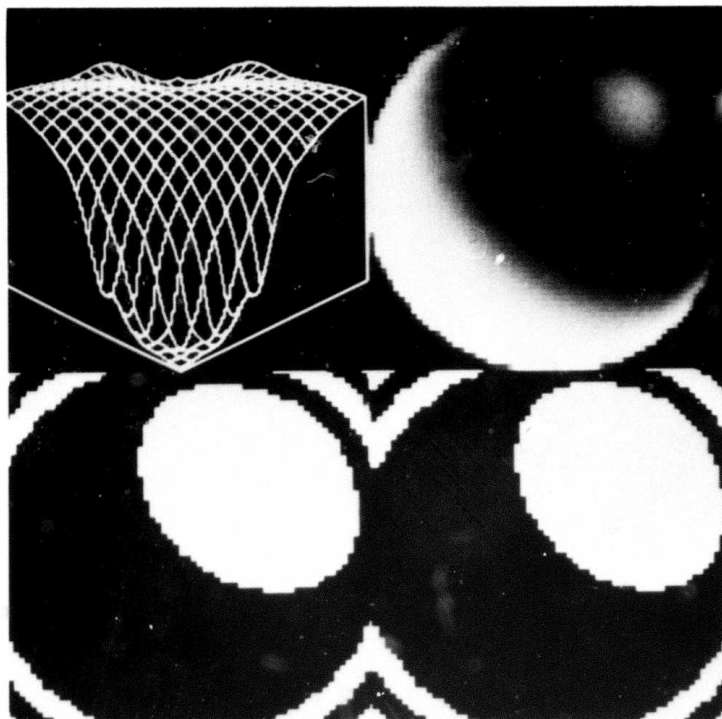


Figure 3. No boundary conditions; planar initial solution perpendicular to viewing direction; image quantization 64×64 .

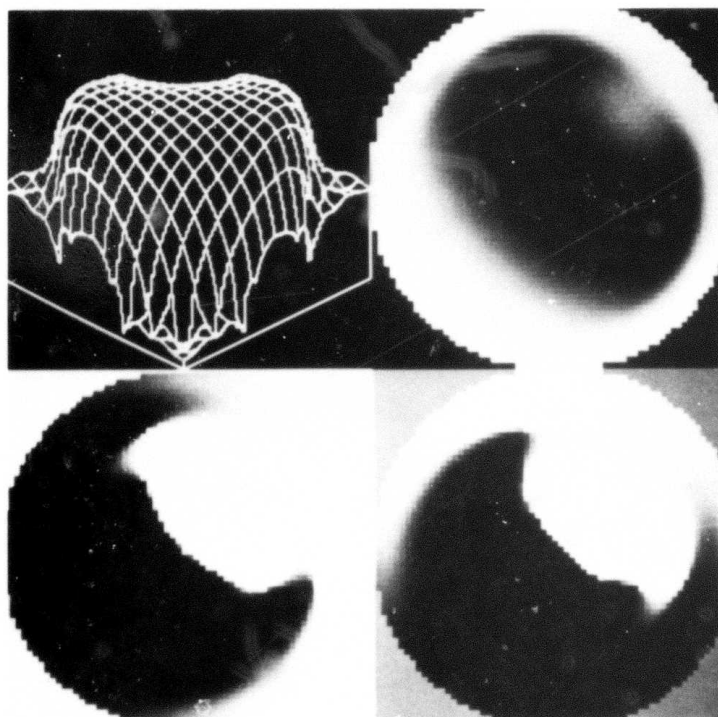


Figure 4. Boundary at edge of sphere given; planar initial solution perpendicular to viewing direction; image quantization 64×64 .

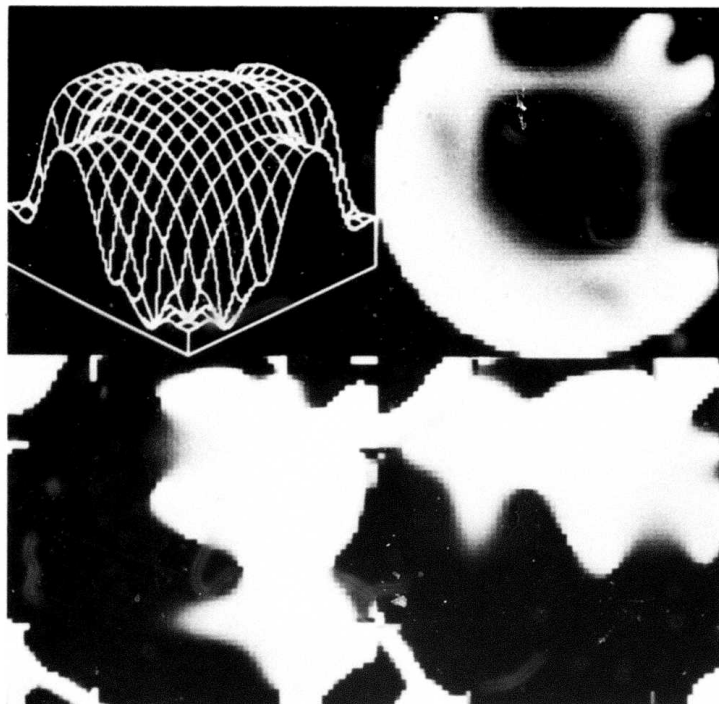


Figure 5. Boundary condition curve on sphere's surface (square shape); planar initial solution perpendicular to viewing direction; image quantization 64×64 .

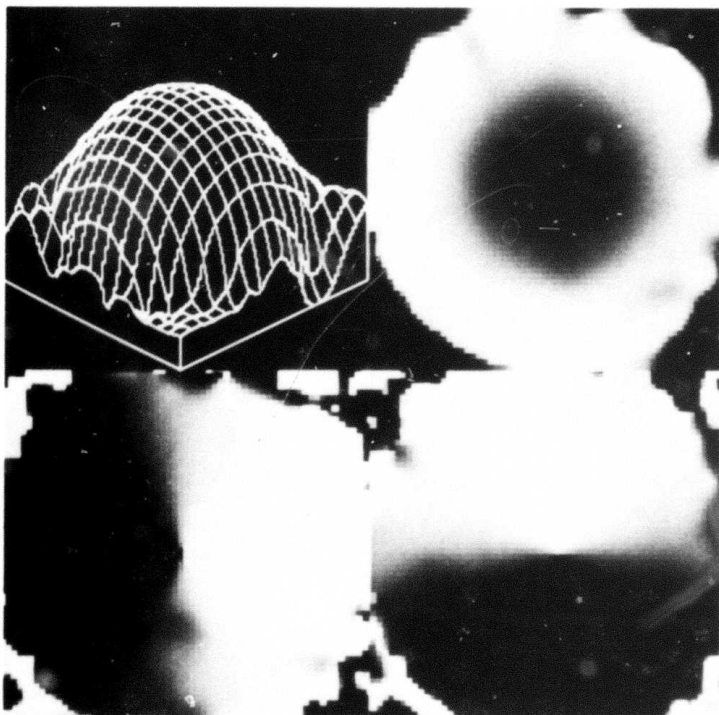


Figure 6. Random five percent of points fixed; planar initial solution perpendicular to viewing direction; image quantization 64×64 .

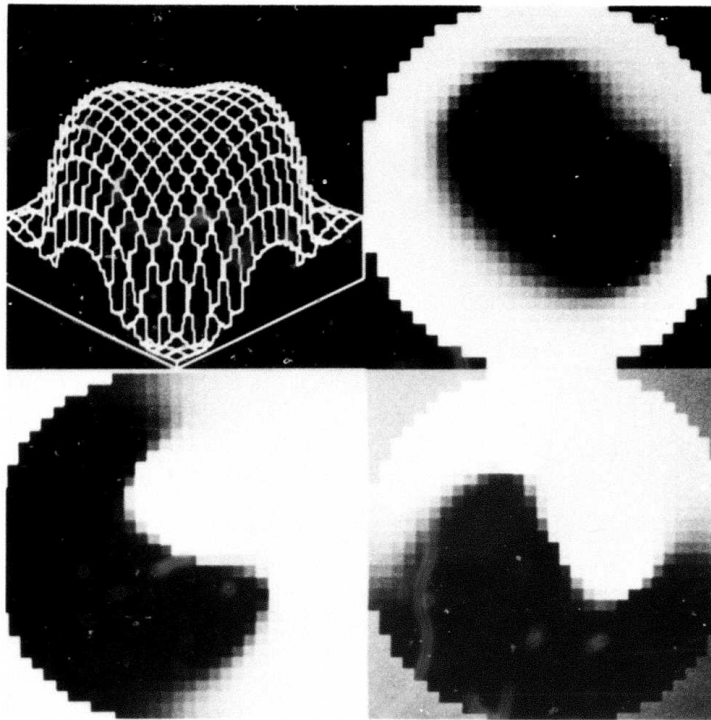


Figure 7. Boundary at edge of sphere given; planar initial solution perpendicular to viewing direction; image quantization 32×32 .

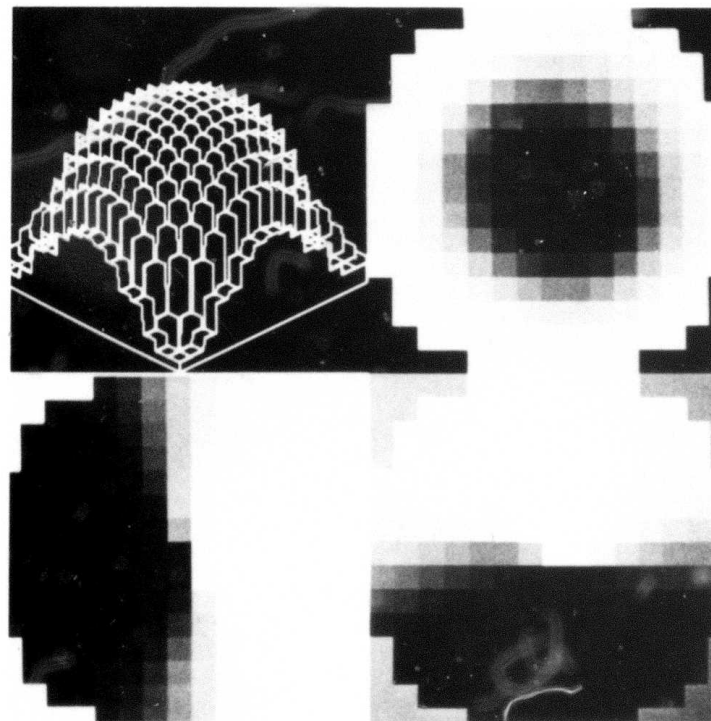


Figure 8. Boundary at edge of sphere given; planar initial solution perpendicular to viewing direction; image quantization 16×16 .

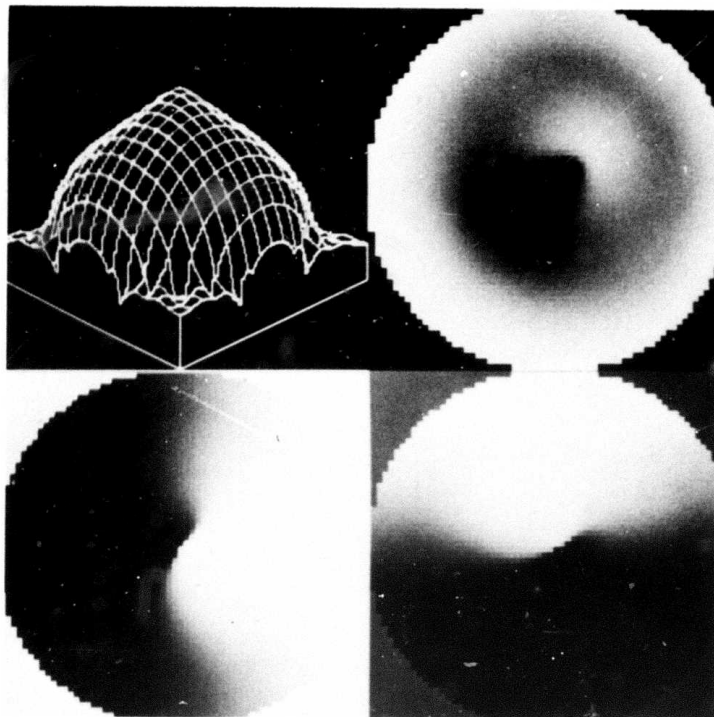


Figure 9. Boundary at edge of sphere given; conical initial solution; image quantization 64×64 .

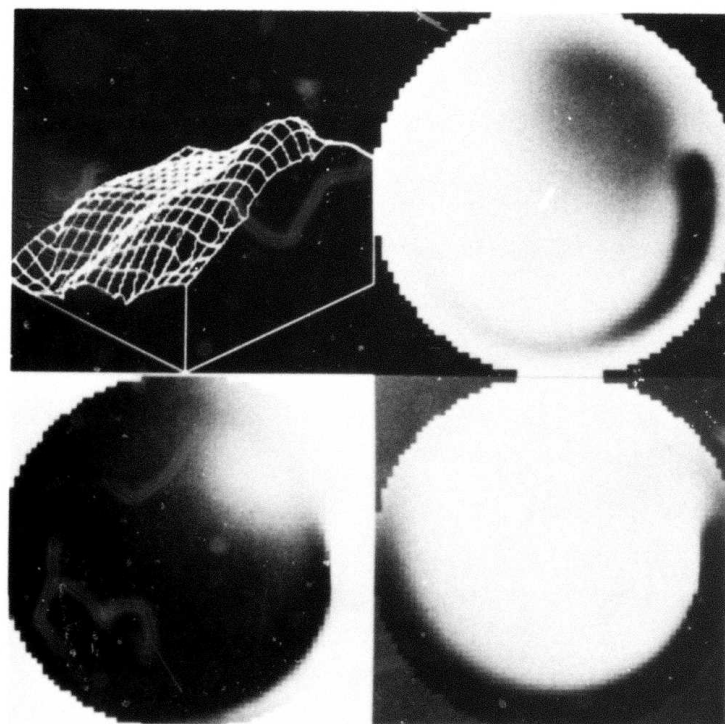


Figure 10. Boundary at edge of sphere given; planar initial solution slanted $\frac{\pi}{4}$ to viewing direction; image quantization 64×64 .

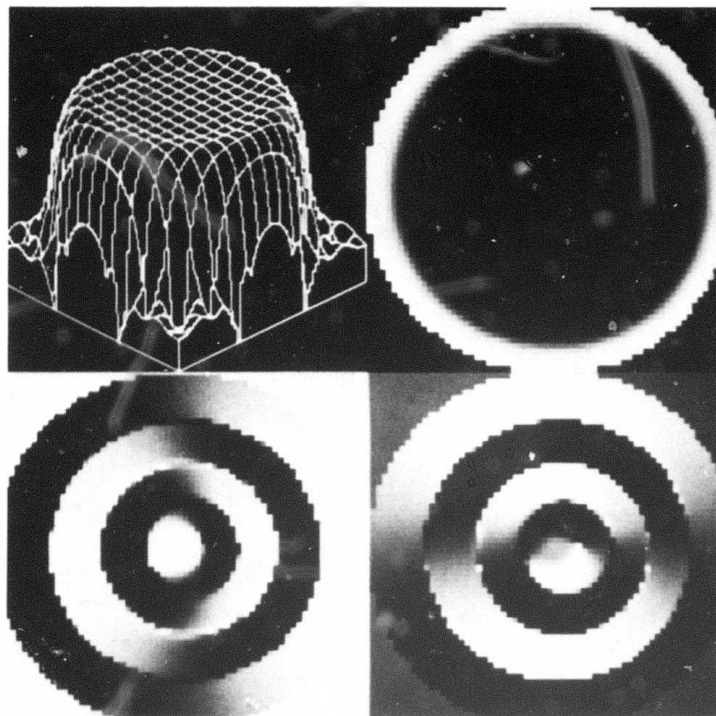


Figure 11. Smoothness constraint only. Boundary at edge of sphere given; planar initial solution perpendicular to viewing direction; image quantization 64×64 .

Appendix C

DETECTION OF THE VISIBLE SKYLINE IN URBAN SCENES AND NATURAL TERRAIN

Michael R. Lowry

I INTRODUCTION

Although not always a well defined problem, delineation of the sky-land boundary provides important constraining information for further analysis of the image. Its very existence in an image tells us something about the location of the camera relative to the scene (i.e. that the scene is being viewed at a high-oblique angle), allows us to estimate visibility, i.e., how far we can see (both as a function of atmospheric viewing conditions, and as a function of the scene content), provides a source of good landmarks for (autonomous) navigation, and defines the boundary beyond which the image no longer depicts portions of the scene having fixed geometric structure.

The objective of this project was to identify the middle-to-far skyline in urban scenes and natural terrain. These two domains span most of the problems encountered in skyline delineation. The principal difficulty is in distinguishing bright clouds from bright objects on the skyline and dark clouds from textureless dark objects on the skyline. Since one of the objectives of identifying the skyline is to constrain further semantic interpretation, the use of detailed semantic models was excluded in this project. Rather the skyline was identified using physical models and assumptions valid for any scene.

One possible use of skyline detection is in landmark identification. Landmark identification has potential use in passive

autonomous navigation and in cross reference and compilation between maps and photographs. Another use is to provide a semantic constraint on further image interpretation. The skyline detection algorithm described in this report is conservative. For images without confusing factors the skyline is identified unambiguously, while in difficult images the picture might be broken up into a sky region, a land region, and a region of ambiguity. Even in these difficult images the algorithm serves to constrain further image interpretation.

Section 2 of this report discusses physical models of sky, cloud, and land luminance and previous work on skyline delineation. Section 3 documents the algorithm developed for skyline delineation: at the core is a region segmentation technique well matched to the physical constraints of the problem. Results are given in Section 4, and the factors influencing the robustness of the algorithm are identified. Section 5 presents our conclusions.

II PHYSICAL MODELS OF SKY, CLOUD, AND LAND LUMINANCE

Different physical mechanisms give rise to the luminance of sky, clouds, and land surfaces. Sky luminance is the result of sunlight scattered by the atmosphere. Cloud luminance is either sunlight attenuated by transmission through a relatively thin cloud or sunlight reflected from a thick cloud. Land and water surfaces reflect sunlight and ambient light.

A. Cloud Free Sky Luminance--Basic Mechanisms

Atmospheric scattering of sunlight occurs through a combination of molecular scattering and aerosol scattering. Molecular scattering, also termed Rayleigh scattering, occurs when the gaseous molecules of the atmosphere interact as dipoles with incident light. The scattering is proportional to the fourth power of the frequency of the incident light. It is the higher scattering of shorter wavelengths which gives the sky its characteristically blue color, and the sun its complementary yellowish tinge. An equal amount of light is scattered forward and backwards along the direction of the incident light. This implies that the sky is equally bright sunside and antisunside.

In contrast to molecular scattering, aerosol scattering is wavelength independent. Aerosol scattering, also termed Mie scattering or large particle scattering, arises from internal refractions and reflection of light that has entered a large particle. Since aerosols scatter all wavelengths equally, the scattered light is white. Thus as the sky gets hazier, it appears whiter. In aerosol scattering there is a much larger forward scattering component than backward scattering component. As a result, on a hazy day the sky is very bright sunside but dark antisunside.

Another key concept in understanding sky luminance is the length of the optical path through the scattering medium from an observer to the effective edge of the atmosphere. Since the density of air decreases by

a factor of 10 for every 15 kilometers of altitude, the scattering properties of the atmosphere effectively terminate at 30-60 kilometers of altitude. As an observer moves his line of sight from the vertical towards the horizon the optical path increases dramatically. Since the optical path passes through proportionally more scattering medium, the sky usually appears brighter towards the horizon. This effect is enhanced because the atmospheric scattering medium near the earth's surface is also being illuminated by light reflected from the land. The exception to this limb effect is when the haze is so dense that the increase in scattering mass is offset by the attenuation of the light passing through the dense aerosol.

The relevant parameters for cloud-free sky luminance are the position of the sun (azimuth and elevation), the aerosol content of the atmosphere, the luminosity of the land, and the line of sight. Quantitative predictions for cloud-free sky luminance were derived through computer simulation by J.V. Dave at IBM Palo Alto [Dave 80]. Five different atmospheric models were used, including aerosol free, aerosol content typical over a land mass, and aerosol content typical over a water mass. Predictions from this study include the following:

- (1) The luminosity of the sky varies gradually and continuously. For an aerosol-free sky the total variation is about 10:1. For an average amount of aerosols there is a strong increase in brightness towards the sun and high noon.
- (2) The clear sky is blue, but becomes whiter with greater aerosol content in the atmosphere.

- (3) An aerosol-free sky is brighter and whiter towards the horizon because of the limb effect. This enhances the contrast at the horizon. However the horizon contrast is attenuated under hazy conditions, particularly for the distant skyline. The horizon contrast is a function of the scattering versus the attenuation properties of the atmosphere.

B. Cloud Luminance--Cumulus and Stratus

The principal difficulty in skyline delineation is distinguishing clouds from land features. Since clouds form a much more limited class than land features, the approach taken in this project has been to use qualitative models of clouds. Good quantitative models of cloud luminosity do not exist. Qualitative predictions can be made from basic theoretical considerations and empirical observations. The basic physical mechanism governing cloud luminosity is aerosol scattering of incident sunlight. For a thin cloud, secondary scattering can be ignored, so the basic mechanism is attenuation of sunlight passing through the cloud. This attenuation obeys Bier's law of transmission--a decaying exponential function of the optical density of the cloud. For the range of optical densities for which secondary scattering can be ignored, the exponential is approximately linear; thus the attenuation is approximately a linear function of the thickness of the cloud. For a thick cloud, the effects of secondary scattering predominate so the cloud is essentially a lambertian surface with 80% reflectivity (hence 20% transmission--there is no adsorption of light in the visible range by water particles). As sunlight passes deeper into a cloud,

successively more and more of the light is scattered in all directions. At some point all the light is scattered, and the illumination is isotropic. Thus 20% of the original illumination is the equilibrium value for isotropic scattering. This lack of directionality in illumination occurs in dense ground fogs and when flying through clouds. Since the sunlit portion of a cumulus cloud acts as a lambertian surface, its shape can be determined from its shading using the same shape-from-shading algorithms used for other lambertian surfaces. In contrast, those cloud surfaces whose luminance is due to transmitted light do not obey the normal surface shape-to-shading relationship.

Thick clouds with large vertical extent are cumulus clouds created by local weather conditions. Local updrafts carry water vapor into the upper atmosphere where it condenses into water droplets. A local downdraft is created by the updraft, providing a natural separation between cumulus clouds. For a cumulus cloud, the top is bright and reflective, the sides not in shadow are similarly bright, and the bottom is dark. Because of the separation between clouds, it is usually possible to see some bright reflective surface. The term "cloudy skies" usually refers to the presence of cumulus clouds.

Stratus clouds are created by larger weather systems, such as low-pressure systems. They have large horizontal extent and comparatively small vertical extent. Consequently from the ground the line of sight is only to the bottom of the cloud, and the luminance is solely from light transmitted through the cloud. Overcast skies are due to stratus clouds.

C. Other Atmospheric Phenomena--Refraction, Reflection, Polarization

The oblique angle of the sun relative to the atmosphere at dawn and dusk causes refraction of incident sunlight according to Snell's laws. The result is a temporal separation of colors as the sun moves towards the horizon. Spectacular atmospheric effects can also be caused by alignment of a large group of refracting or reflecting particles. For example, rainbows are caused by billions of raindrops acting as identical prisms aligned through gravity and the effects of aerodynamic drag. Similarly, ice crystals aligned in a high cloud can act as a gigantic mirror when viewed from an airplane. The effects of refraction and reflection were not incorporated into the implicit assumptions made by the skyline delineation algorithm.

Polarization of the clear sky is due to the interaction of molecular dipoles with incident sunlight. The sun's radiation can be modeled as plane wave oriented perpendicular to the direction of propagation. Within the plane there is no preferred direction of polarization, this isotropy can be represented as a circle in the plane. Incident sunlight causes a molecule to vibrate in sympathy; the light emitted when the molecule vibrates is the scattered light. The vibration of the molecule can be modeled as a circular motion in the plane perpendicular to the sun's rays, since the incident sunlight has no preferred polarization. When this motion is viewed by an observer not directly in line with the sun and the molecule, the motion appears to be tracing out an ellipse. The difference between the major and

minor axis of the ellipse is the degree of polarization. Along any given line of sight, the polarization of the molecular vibration is the same, since the sun is effectively a source of parallel light. In the presence of clouds or haze, polarization is virtually absent since aerosol scattering occurs through refraction and reflection as opposed to the vibration of a molecular dipole. Since it is under conditions of cloud and haze that skyline delineation becomes difficult, polarization is not a disambiguating feature.

D. Use of Color and a World Model in the Recognition of Natural Outdoor Scenes

The doctoral dissertation of Kenneth Sloan was on the use of color and a world model in the recognition of natural outdoor scenes. He used color pictures of wilderness scenes roughly oriented towards the horizon. His objective was to identify regions such as sky, mountains, hills, trees, lakes, and snow patches. The first step in his program was to segment the digitized color picture into regions of homogenous color. The regions of homogenous color were then given semantic labels such as "sky" and merged with nearby regions under control of an extended production system. The productions used contextual and color cues to assign semantic labels and merge regions. The initial goal of the production system was to identify the skyline as a context for further processing.

The opponent process theory of color perception [Hurvich70] [Sloan75] was used to provide color descriptions compatible with human perception. The resulting color descriptions correspond to color names in human language, thereby facilitating extension of the system to man-machine interaction through natural language. The opponent process theory can explain phenomena of human color perception such as color desaturation at low and high intensity, and the effect of surrounding regions on color perception. However these phenomena rely upon the non-linear processing of the initial red/blue/green description. For reasons of computational speed, Sloan implemented the opponent process theory as a simple linear transformation from the initial red/green/blue color space into a yellow-blue/red-green/white-black color space. Thus the only advantage of his implemented opponent process theory was in color description, which was achieved by quantizing each of the three transformed color dimensions into +, -, and 0. This yielded 27 different colors that correspond well to color names in human language.

The initial region segmentation was done by a FORTRAN program which took a seed for each region and recursively found all pixels with 8-neighbor connectivity that were the same color as the initial seed to within a small tolerance. Because of the 8-neighbor connectivity, regions could be very interlocked, as in a jig saw puzzle. Regions were described by area, color of initial seed, bounding rectangle, shape, and texture. Shape and texture were optional descriptions computed at the request of higher-level routines. Shape was described by the distance

from the center of the bounding rectangle to the furthest pixel in each of eight directions. Texture was described by the "busyness" of an area as indicated by the region boundaries.

High-level processing was under the control of an extended production system, which used the contents of Short-Term Memory to focus attention, storing assertions not in current focus under tags in Long-Term Memory. The initial goal was to process the regions in order of decreasing size until some region could be identified as either land or sky. When this was done, the production system was recursively called with a new set of rules for the context established by the identification of the initial region. Regions near the initial region were examined to see if they could be merged into the initial region, or used to establish a new context. For example, if a dark region was found while examining neighbors for a region identified as a patch of sky, then the dark region would be used to establish a context for land and skyline.

The semantic content of the rules used for identification and merging of regions were fairly simple. For example the initial sky region was identified by being near the top of the picture and bright. Tree trunks were identified by being brown and vertically elongated. A nearby region of similar texture and color was merged into a semantically identified region. If one region was above and bluer than an identified region, then it was judged to be further away because of the blue-shift for distant objects. Similar use was made of texture

discontinuities. Dark blue regions were identified as water, while dark blue-green regions are identified as mountains. White regions near a ground patch were identified as snow.

Dr. Sloan's system did not identify the complete skyline in his pictures. Rather it identified fragments of the skyline and depth discontinuities in the land, such as ridgelines, which he called generalized skylines. He used three pictures for the development of his system, and 3 pictures to test the system. His system failed on the picture of Mount Rainier, partially because the glacier was confused with a bank of clouds and partially because the initial region segmentation included regions consisting of both sky and ice. In another picture, of the desolation wilderness, the sky is too dark to be identified as such, but the mountain peaks are identified as depth discontinuities, i.e., as part of the generalized skyline. Other errors include such things as confusing mountain peaks with tree tops and poor identification of bodies of water.

III CURRENT SYSTEM

The current approach is to identify initial sky and land "seeds", that is, regions in the scene that can be confidently identified. The sky seed is classified as being clear, overcast, or cloudy. This information is then used to extend the initial seeds when it can be done with high confidence. The final result is a picture broken into a sky region, a land region, and a possibly empty unknown region.

The picture is assumed to be right side up with the sky dominating the top of the picture and the land filling the entire last row in the bottom of the picture. The current system will almost always find the skyline under conditions of clear sky or under cloudy conditions when the clouds do not touch the skyline nor form an unbroken chain across the sky. This performance is achieved because, for any reasonably sized picture, the luminosity gradient within the sky is always less than the smallest integer-valued threshold possible for region segmentation. Thus the initial sky seed is in fact the entire sky under clear conditions. In conditions of an overcast sky or when cumulus clouds touch the horizon, the system usually finds a bounded region in which the skyline will appear.

The initial region segmentation is done by setting a gray-level contrast threshold and then grouping pixels into 4-neighbor connected regions enclosed by a boundary. The contrast across the boundary

everywhere exceeds the gray-level contrast threshold. This segmentation is accomplished through a one pass scan over the entire picture. A region map is produced, isomorphic to the original picture, with each pixel assigned a region number. Adjacent pixels are given the same region number if their contrast is less than the threshold contrast; otherwise they are assigned different region numbers. During the scan adjacent pixels might be assigned to different regions based upon a local boundary that does not become a closed contour. In this case the two regions are merged; otherwise the algorithm is completely straightforward.

The basic concept behind this method of region segmentation is to combine region-based and edge-based descriptions since edges generally do not form closed contours. An alternative to the approach we employed is to use some edge operator that implicitly computes edge data and region data. The Laplacian operator has been used to derive edge data from zero-crossing contours, it can also be viewed as breaking a picture up into positive and negative regions. However it was empirically determined that the zero-crossing contour corresponding to the skyline was almost inevitably broken at several points, causing the sky region to bleed into the land region. These breaks were usually associated with high-contrast objects near the skyline, and could be quite large in extent. The larger the DOG (difference of Gaussian) operator, the more likely that large objects near the skyline would fall within the same window as the skyline, causing a break in the skyline contour. Smaller

DOG operators were extremely noisy, with many small contours appearing within the sky in addition to the skyline.

Finding region boundaries using gray-level difference thresholds corresponds to edge detection using gradients. Two problems arose, the first being noise causing locally high gradients and the second being large-scale changes escaping detection because of a low gradient. A somewhat less frequent problem was noise acting as a stepping stone between regions and causing them to merge. These problems had a uniform solution: reduce the picture by averaging each $N \times N$ square of pixels into one pixel on a coarser grid. This cancels high-frequency noise because the averaging acts as a low-pass filter. The reduction in size causes large-scale changes to be reduced to changes on the scale of pixel neighbors. Very large-scale changes, such as the gradual change of luminosity over the sky, remain undetected. This is an advantage over region-segmentation techniques that look for regions of constant intensity (or constant color), given that the sky becomes whiter towards the horizon. An added advantage of reducing the picture is that the computation time is reduced by a factor of $N \times N$.

A 4-by-4 picture reduction was chosen arbitrarily. Empirical investigation showed that the region-segmentation technique was fairly robust over a wide range of reduction factors. The constraints are that the reduction average a sufficient number of pixels to cancel out noise and sharpen a hazy skyline.

The skyline-detection algorithm uses a hierarchy of region segmentations. The hierarchy is formed by using five different gray-level difference thresholds for segmentation: 2, 4, 8, 16, 32. Any region found using a threshold of 2 is necessarily a subregion of a region found using a threshold of 4. In general, if T_1 and T_2 are thresholds such that $T_1 < T_2$, then regions segmented using T_1 are subregions of regions segmented using T_2 . A threshold of 2 is the lowest threshold that does not require regions to be of constant intensity. The doubling of thresholds is an arbitrary choice, which provides coverage of edge strengths from 2 to N with only $\log(2) N$ segmentations. It was empirically determined that an upper threshold of 32 was sufficient for a maximum gray-level value of 255.

The initial sky seed is determined with the region segmentation of threshold 2; the region with the largest number of pixels in the top 8% of a picture is chosen as the sky seed. The assumption being made is that some subregion of the true sky is larger than any land subregion in the top 8% of the picture.

The initial land seed is found using the assumption that the highest-contrast continuous boundary spanning the entire horizontal extent of the picture is at or below the skyline. This assumption had only one minor exception in a corpus of 15 pictures used for system development. A bright cumulus cloud on the horizon formed a higher-contrast boundary with the clear sky above it than with the bright rock below it in the picture labeled "Bishop". The method used to find the

initial land seed using this assumption was to grow the initial sky seed until it touched the bottom of the picture. The growth was accomplished by using successively higher thresholds for finding region boundaries. Since the sky is assumed not to touch the bottom of the picture, it is necessary to back off the threshold once the sky region touches the bottom of the picture, this is done by halving the threshold by a factor of 2. The initial land seed is found by finding the transitive closure of all regions connected to the bottom of the picture through neighboring nonsky regions.

The initial sky seed is classified as clear, cloudy, or overcast according to the following criteria. If large regions are enclosed within the initial sky region, the sky is assumed to be cloudy. These large regions usually correspond to bright cumulus clouds, which form high-contrast boundaries. Sometimes the initial sky seed will be a cloud, in which case the interior regions (holes) would be patches of clear sky. A clear sky is approximated well by a linear intensity function over small solid angles. Except for pictures taken with extremely wide-angle lenses, the assumption of a small solid angle is valid. Theoretically the picture should also lie entirely within one quadrant of the sky with respect to the position of the sun, since there is usually an inflection point in the luminosity of the sky at the azimuth and elevation of the sun. In practice the latter criteria does not make a difference, so the intensity of the clear sky is well modeled as a linear function of the x and y coordinates. An overcast sky is

illuminated by sunlight transmitted through layers of clouds. The optical thickness of the clouds determines the luminosity. The optical thickness is usually not well approximated by a linear function over the viewing angle of the picture. The initial sky seed is classified as clear if the mean square error of the least squares linear approximation is less than the mean pixel noise. The pixel noise was computed assuming that the gray-level differences between adjacent pixels in the picture were predominantly due to noise. A histogram was made of adjacent pixel gray-level differences, and the 66% percentile was assumed to represent the standard deviation in an underlying gaussian distribution. This gives a good upper bound when noise is due to the digitization process.

IV RESULTS

The skyline delineation algorithm was tested on a corpus of 15 pictures. These pictures were deliberately selected at the beginning of the project to present difficulties to any mechanical system. The corpus can be classified as follows:

Clear Sky : 2
Haze : 2
Cloudy : 6
Overcast : 5

A. Confusing Factors

In four pictures there is contrast reversal at the horizon, that is objects on the horizon are brighter than the sky above them. In seven of the pictures there are portions of the sky with lower intensity values than some objects on the horizon. In five of the pictures some boundaries within the sky form a higher contrast than the skyline boundary.

B. Performance of Skyline Delineation Algorithm

Clear Sky: In these cases the program found the skyline with no ambiguity. The region segmentation is well matched to the gradual change of intensity over the clear sky, and the brightening of the sky at the horizon increases the skyline contrast, usually making it the highest-contrast boundary in the picture.

Haze: In one case the initial sky seed found in the picture was identical to the sky, while about half the land was declared ambiguous. In the other case the program mistakenly identified the near skyline as the real skyline. The picture was taken from the top of a mountain with Vancouver Bay in the background of a ridgeline. The bay, and the peninsula beyond, are so hazy that the program misidentifies them as a cloud (as do most humans). In many hazy pictures the horizon is so indistinct that the best any interpreter could do, mechanical or human, would be to identify some region of ambiguity in which the skyline must fall.

Cloudy: In none of these pictures is the skyline found unambiguously. In three of these six pictures the sky portion and the land portion of the picture touch at some points. In one of these pictures a small error is made where a cumulus cloud is included as part of the land seed. The cumulus cloud has a very similar intensity to the bright rock beneath it, as opposed to the darker sky above it. In one of these pictures a cloud/snow boundary has no contrast.

Overcast: In one of the overcast pictures, the skyline was identified without ambiguity. In the remaining four pictures, the ambiguous region extended over various portions of the picture. With an overcast picture the land seed is apt to cover most of the actual land in the picture, while the sky is broken at horizontal layers. Thus, in most of the pictures, the ambiguous portion is largely within the sky region

V CONCLUSIONS

The principal difficulty in image understanding is to translate a given interpretation problem into one of identifying observable features of an image. This requires an analysis of the underlying physical mechanisms governing the visual properties of the objects of interest and the image-acquisition process. The analysis yields constraints for translating from observable features back into semantic objects. When

the problem is completely constrained the translation is unambiguous, otherwise ambiguity remains.

This analysis was carried out for the problem of differentiating sky from land. The analysis principally focused on the question of how to extend a point identified as sky into as large a portion of the true sky without misidentifying some portion of land as part of the sky. The principal observable due to physical mechanisms of sky luminance is gradual change. Unlike other problems where geometric structure is transformed into observables through the imaging process, the sky has no geometric structure and there is no fixed model of the geometric structure in the land. The geometric structure which is assumed is that the picture is taken right side up with the sky dominating the top of the picture and not touching the bottom of the picture. This presupposition is equivalent to requiring the question of skyline delineation to be meaningful for the images analyzed.

An algorithm (a region-based segmentation process that splits regions at discontinuities) based upon these constraints was constructed and tested. The algorithm incorporates sufficient constraints to find the skyline unambiguously in easy cases, while delineating an ambiguous region in more difficult cases. The only major mistakes it made were in cases of extreme haze, where humans would also make mistakes.

REFERENCES

- [Dave80] Dave, V.J. (September 1980) "Atmospheric Modeling - Transfer of Visible Radiation in the Atmosphere", Palo Alto Scientific Center, IBM Report No. 6320-3411. To be published in Atmospheric Environment.
- [Hurvich70] Hurvich, Leo M., and Jameson, Dorothea. "Color Vision and color Coding", in PERCEPTION AND ITS DISORDERS Res. Publ. A.R.N.M.D. Vol. XLVIII, The Association for Research in Nervous and Mental Disease.
- [Sloan75] Sloan, Kenneth K. and Bajcsy, Ruzena. "A Computational Structure for Color Perception", ACM 1975 Proceedings. pp 42-45.
- [Sloan77] Sloan, Kenneth. "WORLD MODEL DRIVEN RECOGNITION of NATURAL SCENES", Ph^D Thesis, 1977, University of Pennsylvania, The Moore School of Electrical Engineering, Philadelphia, Pennsylvania 19104.



FIGURE 1 URBAN SCENE WITH OVERCAST SKY

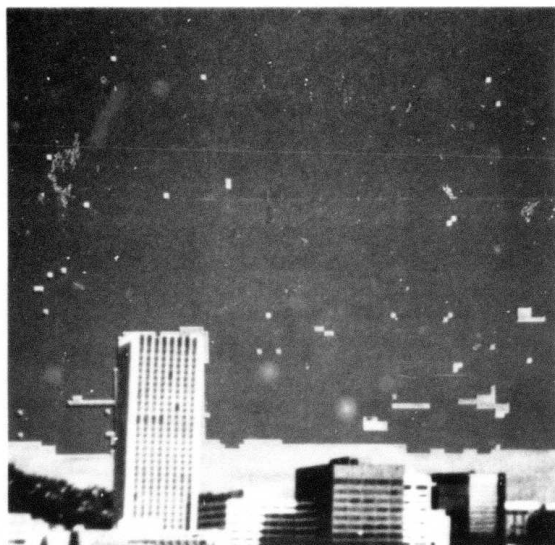


FIGURE 2 SKY SEED FOUND WITH REGION SEGMENTATION AT THRESHOLD 2

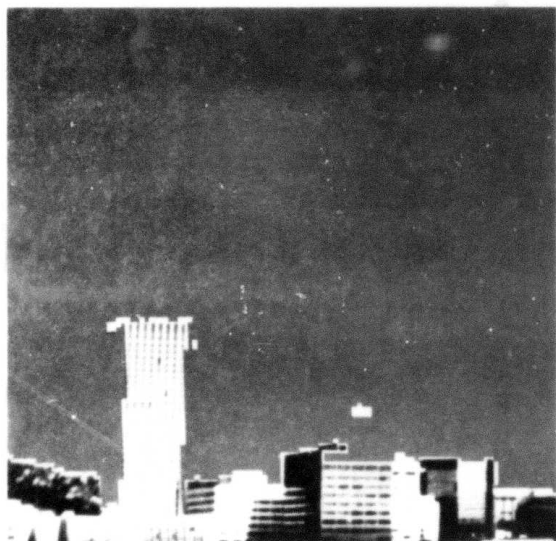


FIGURE 3 THRESHOLD IS DOUBLED UNTIL REGION CONTAINING SKY SEED ALMOST TOUCHES "BOTTOM" OF THE PICTURE. THE LAND SEED IS THE COMPLEMENT.

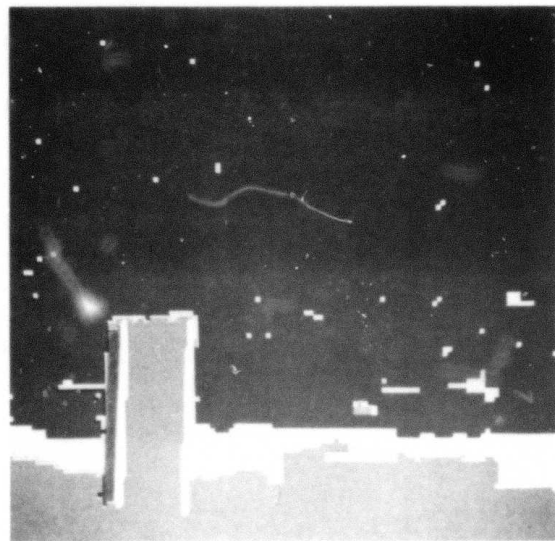
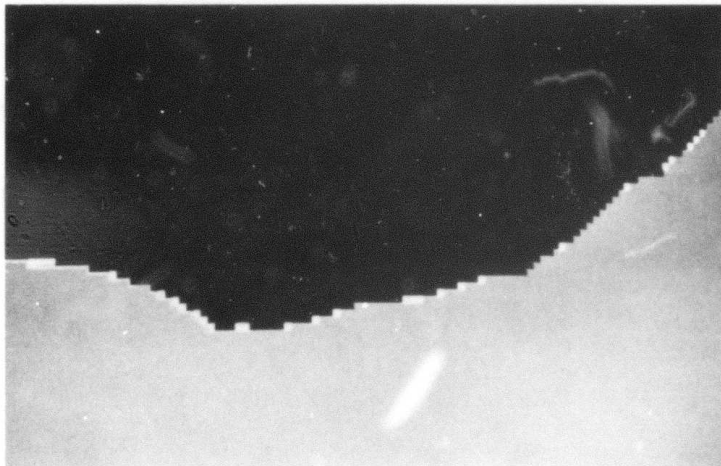


FIGURE 4 PICTURE SEGMENTED INTO SKY SEED, UNCLASSIFIED PORTION, AND LAND SEED



(a) ORIGINAL IMAGE



(b) SKY AND LAND SEED BOUNDARIES COINCIDE AT SKYLINE TO PRODUCE UNAMBIGUOUS DELINEATION

FIGURE 5 SKYLINE DELINEATION